Please type a plus sign (+) inside this box [+]

PTO/SB/05 (2/98)

UTILITY PATENT APPLICATION TRANSMITTAL

(Only for nonprovisional applications under 37 CFR § 1 53(b))

Attorney Docket No	210121.471C1 BOX. SEQ
First Inventor or Application Identifier	Jiangchun Xu
Title	COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS OF COLON CANCER AND METHODS FOR THEIR USE
Express Mail Label No	EL255095431US

APPLICATION ELEMENTS

See MPEP chapter 600 concerning utility patent application contents

ADDRESS TO: Box Patent Application, Assistant Commissioner for Patents, Washington, D.C. 20231

1. [ ] General Authorization Form & Fee Transmittal *(Submit an original and a duplicate for fee processing)*
2. [X] Specification [Total Pages] 69 *(preferred arrangement set forth below)*
   - Descriptive Title of the Invention
   - Cross References to Related Applications
   - Statement Regarding Fed sponsored R & D
   - Reference to Microfiche Appendix
   - Background of the Invention
   - Brief Summary of the Invention
   - Brief Description of the Drawings *(if filed)*
   - Detailed Description
   - Claim(s)
   - Abstract of the Disclosure
3. [X] Drawing(s) *(35 USC 113)* [Total Sheets] 4
4. Oath or Declaration [Total Pages]
   a. [ ] Newly executed (original or copy)
   b. [ ] Copy from a prior application (37 CFR 1 63(d)) *(for continuation/divisional with Box 17 completed)*
      i. [ ] DELETION OF INVENTOR(S) Signed statement attached deleting inventor(s) named in the prior application, see 37 CFR 1.63(d)(2) and 1 33(b)
5. [ ] Incorporation By Reference *(useable if box 4b is checked)* The entire disclosure of the prior application, from which a copy of the oath or declaration is supplied under Box 4b, is considered to be part of the disclosure of the accompanying application and is hereby incorporated by reference therein
6. [ ] Microfiche Computer Program *(Appendix)*
7. Nucleotide and Amino Acid Sequence Submission *(if applicable, all necessary)*
   a. [X] Computer-Readable Copy
   b [X] Paper Copy (identical to computer copy)
   c. [X] Statement verifying identity of above copies

ACCOMPANYING APPLICATION PARTS

8. [ ] Assignment Papers (cover sheet & document(s))
9. [ ] 37 CFR 3.73(b) Statement *(when there is an assignee)* [ ] Power of Attorney
10. [ ] English Translation Document *(if applicable)*
11. [ ] Information Disclosure Statement (IDS)/PTO-1449 [ ] Copies of IDS Citations
12. [ ] Preliminary Amendment
13. [X] Return Receipt Postcard
14. [ ] Small Entity Statement(s) [ ] Statement filed in prior application Status still proper and desired
15. [ ] Certified Copy of Priority Document(s) *(if foreign priority is claimed)*
16. [X] Other: Certificate of Express Mail

17. **If a CONTINUING APPLICATION,** *check appropriate box and supply the requisite information below and in a preliminary amendment*

[ ] Continuation [ ] Divisional [X] Continuation-In-Part (CIP) of prior Application No.: 09/221,298

Prior application information Examiner _____ Group / Art Unit _____

[ ] Claims the benefit of Provisional Application No _____

CORRESPONDENCE ADDRESS

David J. Maki
Seed and Berry LLP
6300 Columbia Center
701 Fifth Avenue
Seattle, Washington 98104-7092
(206) 622-4900 *phone*
(206) 682-6031 *fax*

Respectfully submitted,

TYPED or PRINTED NAME David J. Maki REGISTRATION NO. 31,392

SIGNATURE _____ Date July 2, 1999

C:\CORIXA\3315.DOC

COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS OF COLON CANCER AND METHODS FOR THEIR USE

REFERENCE TO RELATED APPLICATIONS

This application is a continuation-in-part of U.S. Patent Application No. 09/221,298, filed December 23, 1998.

TECHNICAL FIELD

The present invention relates generally to therapy and diagnosis of cancer, such as colon cancer. The invention is more specifically related to polypeptides comprising at least a portion of a colon tumor protein, and to polynucleotides encoding such polypeptides. Such polypeptides and polynucleotides may be used in vaccines and pharmaceutical compositions for prevention and treatment of colon cancer, and for the diagnosis and monitoring of such cancers.

BACKGROUND OF THE INVENTION

Cancer is a significant health problem throughout the world. Although advances have been made in detection and therapy of cancer, no vaccine or other universally successful method for prevention or treatment is currently available. Current therapies, which are generally based on a combination of chemotherapy or surgery and radiation, continue to prove inadequate in many patients.

Colon cancer is the second most frequently diagnosed malignancy in the United States as well as the second most common cause of cancer death. An estimated 95,600 new cases of colon cancer will be diagnosed in 1998, with an estimated 47,700 deaths. The five-year survival rate for patients with colorectal cancer detected in an early localized stage is 92%; unfortunately, only 37% of colorectal cancer is diagnosed at this stage. The survival rate drops to 64% if the cancer is allowed to spread to adjacent organs or lymph nodes, and to 7% in patients with distant metastases.

The prognosis of colon cancer is directly related to the degree of penetration of the tumor through the bowel wall and the presence or absence of nodal involvement,

consequently, early detection and treatment are especially important. Currently, diagnosis is aided by the use of screening assays for fecal occult blood, sigmoidoscopy, colonoscopy and double contrast barium enemas. Treatment regimens are determined by the type and stage of the cancer, and include surgery, radiation therapy and/or chemotherapy. Recurrence following surgery (the most common form of therapy) is a major problem and is often the ultimate cause of death. In spite of considerable research into therapies for the disease, colon cancer remains difficult to diagnose and treat.In spite of considerable research into therapies for these and other cancers, colon cancer remains difficult to diagnose and treat effectively. Accordingly, there is a need in the art for improved methods for detecting and treating such cancers. The present invention fulfills these needs and further provides other related advantages.

SUMMARY OF THE INVENTION

Briefly stated, the present invention provides compositions and methods for the diagnosis and therapy of cancer, such as colon cancer. In one aspect, the present invention provides polypeptides comprising at least a portion of a colon tumor protein, or a variant thereof .. Certain portions and other variants are immunogenic, such that the ability of the variant to react with antigen-specific antisera is not substantially diminished. Within certain embodiments, the polypeptide comprises a sequence that is encoded by a polynucleotide sequence selected from the group consisting of: (a) sequences recited in SEQ ID NO: 2, 8, 15, 16, 22, 24, 30, 32-34, 36, 38, 40, 41, 46-49, 52, 54, 59, 60, 65-69, 79, 89, 90, 93, 99-101, 109-111 and 116-119; (b) variants of a sequence recited in SEQ ID NO: 2, 8, 15, 16, 22, 24, 30, 32-34, 36, 38, 40, 41, 46-49, 52, 54, 59, 60, 65-69, 79, 89, 90, 93, 99-101, 109-111 and 116-119; and (c) complements of a sequence of (a) or (b).

The present invention further provides polynucleotides that encode a polypeptide as described above, or a portion thereof (such as a portion encoding at least 15 amino acid residues of a colon tumor protein), expression vectors comprising such polynucleotides and host cells transformed or transfected with such expression vectors.

Within other aspects, the present invention provides pharmaceutical compositions comprising a polypeptide or polynucleotide as described above and a physiologically acceptable carrier.

Within a related aspect of the present invention, vaccines are provided. Such vaccines comprise a polypeptide or polynucleotide as described above and a non-specific immune response enhancer.

The present invention further provides pharmaceutical compositions that comprise: (a) an antibody or antigen-binding fragment thereof that specifically binds to a colon tumor protein; and (b) a physiologically acceptable carrier.

Within further aspects, the present invention provides pharmaceutical compositions comprising: (a) an antigen presenting cell that expresses a polypeptide as described above and (b) a pharmaceutically acceptable carrier or excipient. Antigen presenting cells include dendritic cells, macrophages, monocytes, fibroblasts and B cells.

Within related aspects, vaccines are provided that comprise: (a) an antigen presenting cell that expresses a polypeptide as described above and (b) a non-specific immune response enhancer.

The present invention further provides, in other aspects, fusion proteins that comprise at least one polypeptide as described above, as well as polynucleotides encoding such fusion proteins.

Within related aspects, pharmaceutical compositions comprising a fusion protein, or a polynucleotide encoding a fusion protein, in combination with a physiologically acceptable carrier are provided.

Vaccines are further provided, within other aspects, that comprise a fusion protein, or a polynucleotide encoding a fusion protein, in combination with a non-specific immune response enhancer.

Within further aspects, the present invention provides methods for inhibiting the development of a cancer in a patient, comprising administering to a patient a pharmaceutical composition or vaccine as recited above.

The present invention further provides, within other aspects, methods for removing tumor cells from a biological sample, comprising contacting a biological sample with T cells that specifically react with a colon tumor protein, wherein the step of contacting is performed under conditions and for a time sufficient to permit the removal of cells expressing the protein from the sample.

Within related aspects, methods are provided for inhibiting the development of a cancer in a patient, comprising administering to a patient a biological sample treated as described above.

Methods are further provided, within other aspects, for stimulating and/or expanding T cells specific for a colon tumor protein, comprising contacting T cells with one or more of: (i) a polypeptide as described above; (ii) a polynucleotide encoding such a polypeptide; and/or (iii) an antigen presenting cell that expresses such a polypeptide; under conditions and for a time sufficient to permit the stimulation and/or expansion of T cells. Isolated T cell populations comprising T cells prepared as described above are also provided.

Within further aspects, the present invention provides methods for inhibiting the development of a cancer in a patient, comprising administering to a patient an effective amount of a T cell population as described above.

The present invention further provides methods for inhibiting the development of a cancer in a patient, comprising the steps of: (a) incubating $CD4^+$ and/or $CD8^+$ T cells isolated from a patient with one or more of: (i) a polypeptide comprising at least an immunogenic portion of a colon tumor protein; (ii) a polynucleotide encoding such a polypeptide; and (iii) an antigen-presenting cell that expressed such a polypeptide; and (b) administering to the patient an effective amount of the proliferated T cells, and thereby inhibiting the development of a cancer in the patient. Proliferated cells may, but need not, be cloned prior to administration to the patient.

Within further aspects, the present invention provides methods for determining the presence or absence of a cancer in a patient, comprising: (a) contacting a biological sample obtained from a patient with a binding agent that binds to a polypeptide as recited above; (b) detecting in the sample an amount of polypeptide that binds to the binding agent; and (c) comparing the amount of polypeptide with a predetermined cut-off value, and therefrom determining the presence or absence of a cancer in the patient. Within preferred embodiments, the binding agent is an antibody, more preferably a monoclonal antibody. The cancer may be colon cancer.

The present invention also provides, within other aspects, methods for monitoring the progression of a cancer in a patient. Such methods comprise the steps of: (a) contacting a biological sample obtained from a patient at a first point in time with a binding

agent that binds to a polypeptide as recited above; (b) detecting in the sample an amount of polypeptide that binds to the binding agent; (c) repeating steps (a) and (b) using a biological sample obtained from the patient at a subsequent point in time; and (d) comparing the amount of polypeptide detected in step (c) with the amount detected in step (b) and therefrom monitoring the progression of the cancer in the patient.

The present invention further provides, within other aspects, methods for determining the presence or absence of a cancer in a patient, comprising the steps of: (a) contacting a biological sample obtained from a patient with an oligonucleotide that hybridizes to a polynucleotide that encodes a colon tumor protein; (b) detecting in the sample a level of a polynucleotide, preferably mRNA, that hybridizes to the oligonucleotide; and (c) comparing the level of polynucleotide that hybridizes to the oligonucleotide with a predetermined cut-off value, and therefrom determining the presence or absence of a cancer in the patient. Within certain embodiments, the amount of mRNA is detected via polymerase chain reaction using, for example, at least one oligonucleotide primer that hybridizes to a polynucleotide encoding a polypeptide as recited above, or a complement of such a polynucleotide. Within other embodiments, the amount of mRNA is detected using a hybridization technique, employing an oligonucleotide probe that hybridizes to a polynucleotide that encodes a polypeptide as recited above, or a complement of such a polynucleotide.

In related aspects, methods are provided for monitoring the progression of a cancer in a patient, comprising the steps of: (a) contacting a biological sample obtained from a patient with an oligonucleotide that hybridizes to a polynucleotide that encodes a colon tumor protein; (b) detecting in the sample an amount of a polynucleotide that hybridizes to the oligonucleotide; (c) repeating steps (a) and (b) using a biological sample obtained from the patient at a subsequent point in time; and (d) comparing the amount of polynucleotide detected in step (c) with the amount detected in step (b) and therefrom monitoring the progression of the cancer in the patient.

Within further aspects, the present invention provides antibodies, such as monoclonal antibodies, that bind to a polypeptide as described above, as well as diagnostic kits comprising such antibodies. Diagnostic kits comprising one or more oligonucleotide probes or primers as described above are also provided.

These and other aspects of the present invention will become apparent upon reference to the following detailed description and attached figures. All references disclosed herein are hereby incorporated by reference in their entirety as if each was incorporated individually.

SEQUENCE IDENTIFIERS

SEQ ID NO: 1 is a first determined cDNA sequence for Contig 1, showing homology to Neutrophil Gelatinase Associated Lipocalin.

SEQ ID NO: 2 is the determined cDNA sequence for Contig 2, showing no significant homology to any known genes.

SEQ ID NO: 3 is the determined cDNA sequence for Contig 4, showing homology to Carcinoembryonic antigen.

SEQ ID NO: 4 is the determined cDNA sequence for Contig 5, showing homology to Carcinoembryonic antigen.

SEQ ID NO: 5 is the determined cDNA sequence for Contig 9, showing homology to Carcinoembryonic antigen.

SEQ ID NO: 6 is the determined cDNA sequence for Contig 52, showing homology to Carcinoembryonic antigen.

SEQ ID NO: 7 is the determined cDNA sequence for Contig 6, showing homology toVillin.

SEQ ID NO: 8 is the determined cDNA sequence for Contig 8, showing no significant homology to any known genes.

SEQ ID NO: 9 is the determined cDNA sequence for Contig 10, showing homology to Transforming Growth Factor (BIGH3).

SEQ ID NO: 10 is the determined cDNA sequence for Contig 19, showing homology to Transforming Growth Factor (BIGH3).

SEQ ID NO: 11 is the determined cDNA sequence for Contig 21, showing homology to Transforming Growth Factor (BIGH3).

SEQ ID NO: 12 is the determined cDNA sequence for Contig 11, showing homology to CO-029.

SEQ ID NO: 13 is the determined cDNA sequence for Contig 55, showing homology to CO-029.

SEQ ID NO: 14 is the determined cDNA sequence for Contig 12, showing homology to Chromosome 17, clone hRPC.1171_I_10, also referred to as C798P.

SEQ ID NO: 15 is the determined cDNA sequence for Contig 13, showing no significant homology to any known gene.

SEQ ID NO: 16 is the determined cDNA sequence for Contig 14, also referred to as 14261, showing no significant homology to any known gene.

SEQ ID NO: 17 is the determined cDNA sequence for Contig 15, showing homology to Ets-Related Transcription Factor (ERT).

SEQ ID NO: 18 is the determined cDNA sequence for Contig 16, showing homology to Chromosome 5, PAC clone 228g9 (LBNL H142).

SEQ ID NO: 19 is the determined cDNA sequence for Contig 24, showing homology to Chromosome 5, PAC clone 228g9 (LBNL H142).

SEQ ID NO: 20 is the determined cDNA sequence for Contig 17, showing homology to Cytokeratin.

SEQ ID NO: 21 is the determined cDNA sequence for Contig 18, showing homology to L1-Cadherin.

SEQ ID NO: 22 is the determined cDNA sequence for Contig 20, showing no significant homology to any known gene.

SEQ ID NO: 23 is the determined cDNA sequence for Contig 22, showing homology to Bumetanide-sensitive Na-K-Cl cotransporter (NKCCl).

SEQ ID NO: 24 is the determined cDNA sequence for Contig 23, showing no significant homology to any known gene.

SEQ ID NO: 25 is the determined cDNA sequence for Contig 25, showing homology to Macrophage Inflammatory Protein 3 alpha.

SEQ ID NO: 26 is the determined cDNA sequence for Contig 26, showing homology to Laminin.

SEQ ID NO: 27 is the determined cDNA sequence for Contig 48, showing homology to Laminin.

SEQ ID NO: 28 is the determined cDNA sequence for Contig 27, showing homology to Mytobularin (MTM1).

SEQ ID NO: 29 is the determined cDNA sequence for Contig 28, showing homology to Chromosome 16 BAC clone CIT987SK-A-363E6.

SEQ ID NO: 30 is the determined cDNA sequence for Contig 29, also referred to as C751P and 14247, showing no significant homology to any known gene, but partial homology to Rat GSK-3β-interacting protein Axil homolog.

SEQ ID NO: 31 is the determined cDNA sequence for Contig 30, showing homology to Zinc Finger Transcription Factor (ZNF207).

SEQ ID NO: 32 is the determined cDNA sequence for Contig 31, showing no significant homology to any known gene, but partial homology to Mus musculus GOB-4 homolog.

SEQ ID NO: 33 is the determined cDNA sequence for Contig 35, showing no significant homology to any known gene, but partial homology to Mus musculus GOB-4 homolog.

SEQ ID NO: 34 is the determined cDNA sequence for Contig 32, showing no significant homology to any known gene.

SEQ ID NO: 35 is the determined cDNA sequence for Contig 34, showing homology to Desmoglein 2.

SEQ ID NO: 36 is the determined cDNA sequence for Contig 36, showing no significant homology to any known gene.

SEQ ID NO: 37 is the determined cDNA sequence for Contig 37, showing homology to Putative Transmembrane Protein.

SEQ ID NO: 38 is the determined cDNA sequence for Contig 38, also referred to as C796P and 14219, showing no significant homology to any known gene.

SEQ ID NO: 39 is the determined cDNA sequence for Contig 40, showing homology to Nonspecific Cross-reacting Antigen.

SEQ ID NO: 40 is the determined cDNA sequence for Contig 41, also referred to as C799P and 14308, showing no significant homology to any known gene.

SEQ ID NO: 41 is the determined cDNA sequence for Contig 42, also referred to as C794P and 14309, showing no significant homology to any known gene.

SEQ ID NO: 42 is the determined cDNA sequence for Contig 43, showing homology to Chromosome 1 specific transcript KIAA0487.

SEQ ID NO: 43 is the determined cDNA sequence for Contig 45, showing homology to hMCM2.

SEQ ID NO: 44 is the determined cDNA sequence for Contig 46, showing homology to ETS2.

SEQ ID NO: 45 is the determined cDNA sequence for Contig 49, showing homology to Pump-1.

SEQ ID NO: 46 is the determined cDNA sequence for Contig 50, also referred to as C792P and 18323, showing no significant homology to any known gene.

SEQ ID NO: 47 is the determined cDNA sequence for Contig 51, also referred to as C795P and 14317, showing no significant homology to any known gene.

SEQ ID NO: 48 is the determined cDNA sequence for 11092, showing no significant homology to any known gene.

SEQ ID NO: 49 is the determined cDNA sequence for 11093, showing no significant homology to any known gene.

SEQ ID NO: 50 is the determined cDNA sequence for 11094, showing homology Human Putative Enterocyte Differentiation Protein.

SEQ ID NO: 51 is the determined cDNA sequence for 11095, showing homology to Human Transcriptional Corepressor hKAP1/TIF1B mRNA.

SEQ ID NO: 52 is the determined cDNA sequence for 11096, showing no significant homology to any known gene.

SEQ ID NO: 53 is the determined cDNA sequence for 11097, showing homology to Human Nonspecific Antigen.

SEQ ID NO: 54 is the determined cDNA sequence for 11098, showing no significant homology to any known gene.

SEQ ID NO: 55 is the determined cDNA sequence for 11099, showing homology to Human Pancreatic Secretory Inhibitor (PST) mRNA.

SEQ ID NO: 56 is the determined cDNA sequence for 11186, showing homology to Human Pancreatic Secretory Inhibitor (PST) mRNA.

SEQ ID NO: 57 is the determined cDNA sequence for 11101, showing homology to Human Chromosome X.

SEQ ID NO: 58 is the determined cDNA sequence for 11102, showing homology to Human Chromosome X.

SEQ ID NO: 59 is the determined cDNA sequence for 11103, showing no significant homology to any known gene.

SEQ ID NO: 60 is the determined cDNA sequence for 11174, showing no significant homology to any known gene.

SEQ ID NO: 61 is the determined cDNA sequence for 11104, showing homology to Human mRNA for KIAA0154.

SEQ ID NO: 62 is the determined cDNA sequence for 11105, showing homology toHuman Apurinic/Apyrimidinic Endonuclease (hap1)mRNA.

SEQ ID NO: 63 is the determined cDNA sequence for 11106, showing homology toHuman Chromosome 12p13.

SEQ ID NO: 64 is the determined cDNA sequence for 11107, showing homology to Human 90 kDa Heat Shock Protein.

SEQ ID NO: 65 is the determined cDNA sequence for 11108, showing no significant homology to any known gene.

SEQ ID NO: 66 is the determined cDNA sequence for 11112, showing no significant homology to any known gene.

SEQ ID NO: 67 is the determined cDNA sequence for 11115, showing no significant homology to any known gene.

SEQ ID NO: 68 is the determined cDNA sequence for 11117, showing no significant homology to any known gene.

SEQ ID NO: 69 is the determined cDNA sequence for 11118, showing no significant homology to any known gene.

SEQ ID NO: 70 is the determined cDNA sequence for 11119, showing homology to Human Elongation Factor 1-alpha.

SEQ ID NO: 71 is the determined cDNA sequence for 11121, showing homology to Human Lamin B Receptor (LBR) mRNA.

SEQ ID NO: 72 is the determined cDNA sequence for 11122, showing homology to H. sapiens mRNA for Novel Glucocorticoid.

SEQ ID NO: 73 is the determined cDNA sequence for 11123, showing homology to H. sapiens mRNA for snRNP protein B.

SEQ ID NO: 74 is the determined cDNA sequence for 11124, showing homology to Human Cisplatin Resistance Associated Beta-protein.

SEQ ID NO: 75 is the determined cDNA sequence for 11127, showing homology to M. musculus Calumenin mRNA.

SEQ ID NO: 76 is the determined cDNA sequence for 11128, showing homology to Human ras-related small GTP binding protein.

SEQ ID NO: 77 is the determined cDNA sequence for 11130, showing homology to Human Cosmid U169d2.

SEQ ID NO: 78 is the determined cDNA sequence for 11131, showing homology to H. sapiens mRNA for protein homologous to Elongation 1-g.

SEQ ID NO: 79 is the determined cDNA sequence for 11134, showing no significant homology to any known gene.

SEQ ID NO: 80 is the determined cDNA sequence for 11135, showing homology to H. sapiens Nieman-Pick (NPC1) mRNA.

SEQ ID NO: 81 is the determined cDNA sequence for 11137, showing homology to H. sapiens mRNA for Niecin b-chain.

SEQ ID NO: 82 is the determined cDNA sequence for 11138, showing homology to Human Endogenous Retroviral Protease mRNA.

SEQ ID NO: 83 is the determined cDNA sequence for 11139, showing homology to H. sapiens mRNA for DMBT1 protein.

SEQ ID NO: 84 is the determined cDNA sequence for 11140, showing homology to H. sapiens ras GTPase activating-like protein.

SEQ ID NO: 85 is the determined cDNA sequence for 11143, showing homology to Human Acidic Ribosomal Phosphoprotein PO mRNA.

SEQ ID NO: 86 is the determined cDNA sequence for 11144, showing homology to H. sapiens U21 mRNA.

SEQ ID NO: 87 is the determined cDNA sequence for 11145, showing homology to Human GTP-binding protein.

SEQ ID NO: 88 is the determined cDNA sequence for 11148, showing homology to H. sapiens U21 mRNA.

SEQ ID NO: 89 is the determined cDNA sequence for 11151, showing no significant homology to any known gene.

SEQ ID NO: 90 is the determined cDNA sequence for 11154, showing no significant homology to any known gene.

SEQ ID NO: 91 is the determined cDNA sequence for 11156, showing homology to H. sapiens Ribosomal Protein L27.

SEQ ID NO: 92 is the determined cDNA sequence for 11157, showing homology to H. sapiens Ribosomal Protein L27.

SEQ ID NO: 93 is the determined cDNA sequence for 11158, showing no significant homology to any known gene.

SEQ ID NO: 94 is the determined cDNA sequence for 11162, showing homology to Ag-X antigen.

SEQ ID NO: 95 is the determined cDNA sequence for 11164, showing homology to H. sapiens mRNA for Signal Recognition Protein sub14.

SEQ ID NO: 96 is the determined cDNA sequence for 11165, showing homology to Human PAC 204e5/127h14.

SEQ ID NO: 97 is the determined cDNA sequence for 11166, showing homology to Human mRNA for KIAA0108.

SEQ ID NO: 98 is the determined cDNA sequence for 11167, showing homology to H. sapiens mRNA for Neutrophil Gelatinase assct. Lipocalin.

SEQ ID NO: 99 is the determined cDNA sequence for 11168, showing no significant homology to any known gene.

SEQ ID NO: 100 is the determined cDNA sequence for 11172, showing no significant homology to any known gene.

SEQ ID NO: 101 is the determined cDNA sequence for 11175, showing no significant homology to any known gene.

SEQ ID NO: 102 is the determined cDNA sequence for 11176, showing homology to Human maspin mRNA.

SEQ ID NO: 103 is the determined cDNA sequence for 11177, showing homology to Human Carcinoembryonic Antigen.

SEQ ID NO: 104 is the determined cDNA sequence for 11178, showing homology to Human A-Tubulin mRNA.

SEQ ID NO: 105 is the determined cDNA sequence for 11179, showing homology to Human mRNA for proton-ATPase-like protein.

SEQ ID NO: 106 is the determined cDNA sequence for 11180, showing homology to Human HepG2 3' region cDNA clone hmd.

SEQ ID NO: 107 is the determined cDNA sequence for 11182, showing homology to Human MHC homologous to Chicken B-Complex Protein.

SEQ ID NO: 108 is the determined cDNA sequence for 11183, showing homology to Human High Mobility Group Box (SSRP1) mRNA.

SEQ ID NO: 109 is the determined cDNA sequence for 11184, showing no significant homology to any known gene.

SEQ ID NO: 110 is the determined cDNA sequence for 11185, showing no significant homology to any known gene.

SEQ ID NO: 111 is the determined cDNA sequence for 11187, showing no significant homology to any known gene.

SEQ ID NO: 112 is the determined cDNA sequence for 11190, showing homology to Human Replication Protein A 70kDa.

SEQ ID NO: 113 is the determined cDNA sequence for Contig 47, also referred to as C797P, showing homology to Human Chromosome X clone bWXD342.

SEQ ID NO: 114 is the determined cDNA sequence for Contig 7, showing homology to Equilibrative Nucleoside Transporter 2 (ent2).

SEQ ID NO: 115 is the determined cDNA sequence for 14235.1, also referred to as C791P, showing homology to H. sapiens chromosome 21 derived BAC containing ets-2 gene.

SEQ ID NO: 116 is the determined cDNA sequence for 14287.2, showing no significant homology to any known gene, but some degree of homology to Putative Transmembrane Protein.

SEQ ID NO: 117 is the determined cDNA sequence for 14233.1, also referred to as Contig 48, showing no significant homology to any known gene.

SEQ ID NO: 118 is the determined cDNA sequence for 14298.2, also referred to as C793P, showing no significant homology to any known gene.

SEQ ID NO: 119 is the determined cDNA sequence for 14372, also referred to as Contig 44, showing no significant homology to any known gene.

SEQ ID NO: 120 is the determined cDNA sequence for 14372, also referred to as 14295, showing homology to secreted cement gland protein XAG-2 homolog.

BRIEF DESCRIPTION OF THE DRAWINGS

Figures 1A-D are determined full length cDNA sequence and corresponding amino acid sequence for P733P, which correspond with the known gene Beta IG-H3.

DETAILED DESCRIPTION OF THE INVENTION

As noted above, the present invention is generally directed to compositions and methods for the therapy and diagnosis of cancer, such as colon cancer. The compositions described herein may include colon tumor polypeptides, polynucleotides encoding such polypeptides, binding agents such as antibodies, antigen presenting cells (APCs) and/or immune system cells (*e.g.,* T cells). Polypeptides of the present invention generally comprise at least a portion (such as an immunogenic portion) of a colon tumor protein or a variant thereof. A "colon tumor protein" is a protein that is expressed in colon tumor cells at a level that is at least two fold, and preferably at least five fold, greater than the level of expression in a normal tissue, as determined using a representative assay provided herein . Certain colon tumor proteins are tumor proteins that react detectably (within an immunoassay, such as an

ELISA or Western blot) with antisera of a patient afflicted with colon cancer. Polynucleotides of the subject invention generally comprise a DNA or RNA sequence that encodes all or a portion of such a polypeptide, or that is complementary to such a sequence. Antibodies are generally immune system proteins, or antigen-binding fragments thereof, that are capable of binding to a polypeptide as described above. Antigen presenting cells include dendritic cells,macrophages, monocytes, fibroblasts and B-cells that express a polypeptide as described above. T cells that may be employed within such compositions are generally T cells that are specific for a polypeptide as described above.

The present invention is based on the discovery of human colon tumor proteins. Partial sequences of polynucleotides encoding specific tumor proteins are provided in SEQ ID NOs: 1-120.

COLON TUMOR PROTEIN POLYNUCLEOTIDES

Any polynucleotide that encodes a colon tumor protein or a portion or other variant thereof as described herein is encompassed by the present invention. Preferred polynucleotides comprise at least 15 consecutive nucleotides, preferably at least 30 consecutive nucleotides and more preferably at least 45 consecutive nucleotides, that encode a portion of a colon tumor protein. More preferably, a polynucleotide encodes an immunogenic portion of a colon tumor protein. Polynucleotides complementary to any such sequences are also encompassed by the present invention. Polynucleotides may be single-stranded (coding or antisense) or double-stranded, and may be DNA (genomic, cDNA or synthetic) or RNA molecules. RNA molecules include HnRNA molecules, which contain introns and correspond to a DNA molecule in a one-to-one manner, and mRNA molecules, which do not contain introns. Additional coding or non-coding sequences may, but need not, be present within a polynucleotide of the present invention, and a polynucleotide may, but need not, be linked to other molecules and/or support materials.

Polynucleotides may comprise a native sequence (*i.e.*, an endogenous sequence that encodes a colon tumor protein or a portion thereof) or may comprise a variant of such a sequence. Polynucleotide variants may contain one or more substitutions, additions, deletions and/or insertions such that the immunogenicity of the encoded polypeptide is not diminished, relative to a native tumor protein. The effect on the

immunogenicity of the encoded polypeptide may generally be assessed as described herein. Variants preferably exhibit at least about 70% identity, more preferably at least about 80% identity and most preferably at least about 90% identity to a polynucleotide sequence that encodes a native colon tumor protein or a portion thereof.

Two polynucleotide or polypeptide sequences are said to be "identical" if the sequence of nucleotides or amino acids in the two sequences is the same when aligned for maximum correspondence as described below. Comparisons between two sequences are typically performed by comparing the sequences over a comparison window to identify and compare local regions of sequence similarity. A "comparison window" as used herein, refers to a segment of at least about 20 contiguous positions, usually 30 to about 75, 40 to about 50, in which a sequence may be compared to a reference sequence of the same number of contiguous positions after the two sequences are optimally aligned.

Optimal alignment of sequences for comparison may be conducted using the Megalign program in the Lasergene suite of bioinformatics software (DNASTAR, Inc., Madison, WI), using default parameters. This program embodies several alignment schemes described in the following references: Dayhoff, M.O. (1978) A model of evolutionary change in proteins – Matrices for detecting distant relationships. In Dayhoff, M.O. (ed.) Atlas of Protein Sequence and Structure, National Biomedical Research Foundation, Washington DC Vol. 5, Suppl. 3, pp. 345-358; Hein J. (1990) Unified Approach to Alignment and Phylogenes pp. 626-645 *Methods in Enzymology* vol. 183, Academic Press, Inc., San Diego, CA; Higgins, D.G. and Sharp, P.M. (1989) *CABIOS 5*:151-153; Myers, E.W. and Muller W. (1988) *CABIOS 4*:11-17; Robinson, E.D. (1971) *Comb. Theor 11*:105; Santou, N. Nes, M. (1987) *Mol. Biol. Evol. 4*:406-425; Sneath, P.H.A. and Sokal, R.R. (1973) *Numerical Taxonomy – the Principles and Practice of Numerical Taxonomy*, Freeman Press, San Francisco, CA; Wilbur, W.J. and Lipman, D.J. (1983) *Proc. Natl. Acad., Sci. USA 80*:726-730.

Preferably, the "percentage of sequence identity" is determined by comparing two optimally aligned sequences over a window of comparison of at least 20 positions, wherein the portion of the polynucleotide or polypeptide sequence in the comparison window may comprise additions or deletions (i.e. gaps) of 20 percent or less, usually 5 to 15 percent, or 10 to 12 percent, as compared to the reference sequences (which does not comprise

additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which the identical nucleic acid bases or amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the reference sequence (i.e. the window size) and multiplying the results by 100 to yield the percentage of sequence identity.

Variants may also, or alternatively, be substantially homologous to a native gene, or a portion or complement thereof. Such polynucleotide variants are capable of hybridizing under moderately stringent conditions to a naturally occurring DNA sequence encoding a native colon tumor protein (or a complementary sequence). Suitable moderately stringent conditions include prewashing in a solution of 5 X SSC, 0.5% SDS, 1.0 mM EDTA (pH 8.0); hybridizing at 50°C-65°C, 5 X SSC, overnight; followed by washing twice at 65°C for 20 minutes with each of 2X, 0.5X and 0.2X SSC containing 0.1% SDS.

It will be appreciated by those of ordinary skill in the art that, as a result of the degeneracy of the genetic code, there are many nucleotide sequences that encode a polypeptide as described herein. Some of these polynucleotides bear minimal homology to the nucleotide sequence of any native gene. Nonetheless, polynucleotides that vary due to differences in codon usage are specifically contemplated by the present invention. Further, alleles of the genes comprising the polynucleotide sequences provided herein are within the scope of the present invention. Alleles are endogenous genes that are altered as a result of one or more mutations, such as deletions, additions and/or substitutions of nucleotides. The resulting mRNA and protein may, but need not, have an altered structure or function. Alleles may be identified using standard techniques (such as hybridization, amplification and/or database sequence comparison).

Polynucleotides may be prepared using any of a variety of techniques. For example, a polynucleotide may be identified, as described in more detail below, by screening a microarray of cDNAs for tumor-associated expression (*i.e.*, expression that is at least two fold greater in a colon tumor than in normal tissue , as determined using a representative assay provided herein). Such screens may be performed using an Incyte microarray (Palo Alto, CA) according to the manufacturer's instructions (and essentially as described by Schena et al., *Proc. Natl. Acad. Sci. USA 93*:10614-10619, 1996 and Heller et al., *Proc. Natl.*

Acad. Sci. USA 94:2150-2155, 1997). Alternatively, polypeptides may be amplified from cDNA prepared from cells expressing the proteins described herein, such as colon tumor cells. Such polynucleotides may be amplified via polymerase chain reaction (PCR). For this approach, sequence-specific primers may be designed based on the sequences provided herein, and may be purchased or synthesized.

An amplified portion may be used to isolate a full length gene from a suitable library (*e.g.*, a colon tumor cDNA library) using well known techniques. Within such techniques, a library (cDNA or genomic) is screened using one or more polynucleotide probes or primers suitable for amplification. Preferably, a library is size-selected to include larger molecules. Random primed libraries may also be preferred for identifying 5' and upstream regions of genes. Genomic libraries are preferred for obtaining introns and extending 5' sequences.

For hybridization techniques, a partial sequence may be labeled (*e.g.*, by nick-translation or end-labeling with ^{32}P) using well known techniques. A bacterial or bacteriophage library is then screened by hybridizing filters containing denatured bacterial colonies (or lawns containing phage plaques) with the labeled probe (*see* Sambrook et al., *Molecular Cloning: A Laboratory Manual,* Cold Spring Harbor Laboratories, Cold Spring Harbor, NY, 1989). Hybridizing colonies or plaques are selected and expanded, and the DNA is isolated for further analysis. cDNA clones may be analyzed to determine the amount of additional sequence by, for example, PCR using a primer from the partial sequence and a primer from the vector. Restriction maps and partial sequences may be generated to identify one or more overlapping clones. The complete sequence may then be determined using standard techniques, which may involve generating a series of deletion clones. The resulting overlapping sequences are then assembled into a single contiguous sequence. A full length cDNA molecule can be generated by ligating suitable fragments, using well known techniques.

Alternatively, there are numerous amplification techniques for obtaining a full length coding sequence from a partial cDNA sequence. Within such techniques, amplification is generally performed via PCR. Any of a variety of commercially available kits may be used to perform the amplification step. Primers may be designed using, for example, software well known in the art. Primers are preferably 22-30 nucleotides in length,

have a GC content of at least 50% and anneal to the target sequence at temperatures of about 68°C to 72°C. The amplified region may be sequenced as described above, and overlapping sequences assembled into a contiguous sequence.

One such amplification technique is inverse PCR (*see* Triglia et al., *Nucl. Acids Res. 16*:8186, 1988), which uses restriction enzymes to generate a fragment in the known region of the gene. The fragment is then circularized by intramolecular ligation and used as a template for PCR with divergent primers derived from the known region. Within an alternative approach, sequences adjacent to a partial sequence may be retrieved by amplification with a primer to a linker sequence and a primer specific to a known region. The amplified sequences are typically subjected to a second round of amplification with the same linker primer and a second primer specific to the known region. A variation on this procedure, which employs two primers that initiate extension in opposite directions from the known sequence, is described in WO 96/38591. Another such technique is known as "rapid amplification of cDNA ends" or RACE. This technique involves the use of an internal primer and an external primer, which hybridizes to a polyA region or vector sequence, to identify sequences that are 5' and 3' of a known sequence. Additional techniques include capture PCR (Lagerstrom et al., *PCR Methods Applic. 1*:111-19, 1991) and walking PCR (Parker et al., *Nucl. Acids. Res. 19*:3055-60, 1991). Other methods employing amplification may also be employed to obtain a full length cDNA sequence.

In certain instances, it is possible to obtain a full length cDNA sequence by analysis of sequences provided in an expressed sequence tag (EST) database, such as that available from GenBank. Searches for overlapping ESTs may generally be performed using well known programs (*e.g.,* NCBI BLAST searches), and such ESTs may be used to generate a contiguous full length sequence.

Certain nucleic acid sequences of cDNA molecules encoding portions of colon tumor proteins are provided in SEQ ID NO: 1-120. These polynucleotides were isolated from colon tumor cDNA libraries using conventional and/or PCR-based subtraction techniques, as described below.

Polynucleotide variants may generally be prepared by any method known in the art, including chemical synthesis by, for example, solid phase phosphoramidite chemical synthesis. Modifications in a polynucleotide sequence may also be introduced using standard

mutagenesis techniques, such as oligonucleotide-directed site-specific mutagenesis (*see* Adelman et al., *DNA 2*:183, 1983). Alternatively, RNA molecules may be generated by *in vitro* or *in vivo* transcription of DNA sequences encoding a colon tumor protein, or portion thereof, provided that the DNA is incorporated into a vector with a suitable RNA polymerase promoter (such as T7 or SP6). Certain portions may be used to prepare an encoded polypeptide, as described herein. In addition, or alternatively, a portion may be administered to a patient such that the encoded polypeptide is generated *in vivo* (*e.g.*, by transfecting antigen-presenting cells, such as dendritic cells, with a cDNA construct encoding a colon tumor polypeptide, and administering the transfected cells to the patient).

A portion of a sequence complementary to a coding sequence (*i.e.*, an antisense polynucleotide) may also be used as a probe or to modulate gene expression. cDNA constructs that can be transcribed into antisense RNA may also be introduced into cells of tissues to facilitate the production of antisense RNA. An antisense polynucleotide may be used, as described herein, to inhibit expression of a tumor protein. Antisense technology can be used to control gene expression through triple-helix formation, which compromises the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors or regulatory molecules (*see* Gee et al., *In* Huber and Carr, *Molecular and Immunologic Approaches*, Futura Publishing Co. (Mt. Kisco, NY; 1994)). Alternatively, an antisense molecule may be designed to hybridize with a control region of a gene (*e.g.*. promoter, enhancer or transcription initiation site), and block transcription of the gene; or to block translation by inhibiting binding of a transcript to ribosomes.

A portion of a coding sequence, or of a complementary sequence, may also be designed as a probe or primer to detect gene expression. Probes may be labeled with a variety of reporter groups, such as radionuclides and enzymes, and are preferably at least 10 nucleotides in length, more preferably at least 20 nucleotides in length and still more preferably at least 30 nucleotides in length. Primers, as noted above, are preferably 22-30 nucleotides in length.

Any polynucleotide may be further modified to increase stability *in vivo*. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends; the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages in the backbone; and/or the inclusion of nontraditional bases such

as inosine, queosine and wybutosine, as well as acetyl- methyl-, thio- and other modified forms of adenine, cytidine, guanine, thymine and uridine.

Nucleotide sequences as described herein may be joined to a variety of other nucleotide sequences using established recombinant DNA techniques. For example, a polynucleotide may be cloned into any of a variety of cloning vectors, including plasmids, phagemids, lambda phage derivatives and cosmids. Vectors of particular interest include expression vectors, replication vectors, probe generation vectors and sequencing vectors. In general, a vector will contain an origin of replication functional in at least one organism, convenient restriction endonuclease sites and one or more selectable markers. Other elements will depend upon the desired use, and will be apparent to those of ordinary skill in the art.

Within certain embodiments, polynucleotides may be formulated so as to permit entry into a cell of a mammal, and expression therein. Such formulations are particularly useful for therapeutic purposes, as described below. Those of ordinary skill in the art will appreciate that there are many ways to achieve expression of a polynucleotide in a target cell, and any suitable method may be employed. For example, a polynucleotide may be incorporated into a viral vector such as, but not limited to, adenovirus, adeno-associated virus, retrovirus, or vaccinia or other pox virus (*e.g.*, avian pox virus). Techniques for incorporating DNA into such vectors are well known to those of ordinary skill in the art. A retroviral vector may additionally transfer or incorporate a gene for a selectable marker (to aid in the identification or selection of transduced cells) and/or a targeting moiety, such as a gene that encodes a ligand for a receptor on a specific target cell, to render the vector target specific. Targeting may also be accomplished using an antibody, by methods known to those of ordinary skill in the art.

Other formulations for therapeutic purposes include colloidal dispersion systems, such as macromolecule complexes, nanocapsules, microspheres, beads, and lipid-based systems including oil-in-water emulsions, micelles, mixed micelles, and liposomes. A preferred colloidal system for use as a delivery vehicle *in vitro* and *in vivo* is a liposome (*i.e.*, an artificial membrane vesicle). The preparation and use of such systems is well known in the art.

COLONTUMOR POLYPEPTIDES

Within the context of the present invention, polypeptides may comprise at least an immunogenic portion of a colon tumor protein or a variant thereof, as described herein. As noted above, a "colon tumor protein" is a protein that is expressed by colon tumor cells. Proteins that are colon tumor proteins also react detectably within an immunoassay (such as an ELISA) with antisera from a patient with colon cancer. Polypeptides as described herein may be of any length. Additional sequences derived from the native protein and/or heterologous sequences may be present, and such sequences may (but need not) possess further immunogenic or antigenic properties.

An "immunogenic portion," as used herein is a portion of a protein that is recognized (*i.e.,* specifically bound) by a B-cell and/or T-cell surface antigen receptor. Such immunogenic portions generally comprise at least 5 amino acid residues, more preferably at least 10, and still more preferably at least 20 amino acid residues of a colon tumor protein or a variant thereof. Certain preferred immunogenic portions include peptides in which an N-terminal leader sequence and/or transmembrane domain have been deleted. Other preferred immunogenic portions may contain a small N- and/or C-terminal deletion (*e.g.,* 1-30 amino acids, preferably 5-15 amino acids), relative to the mature protein.

Immunogenic portions may generally be identified using well known techniques, such as those summarized in Paul, *Fundamental Immunology*, 3rd ed., 243-247 (Raven Press, 1993) and references cited therein. Such techniques include screening polypeptides for the ability to react with antigen-specific antibodies, antisera and/or T-cell lines or clones. As used herein, antisera and antibodies are "antigen-specific" if they specifically bind to an antigen (*i.e.,* they react with the protein in an ELISA or other immunoassay, and do not react detectably with unrelated proteins). Such antisera and antibodies may be prepared as described herein, and using well known techniques. An immunogenic portion of a native colon tumor protein is a portion that reacts with such antisera and/or T-cells at a level that is not substantially less than the reactivity of the full length polypeptide (*e.g.,* in an ELISA and/or T-cell reactivity assay). Such immunogenic portions may react within such assays at a level that is similar to or greater than the reactivity of the full length polypeptide. Such screens may generally be performed using methods well known to those of ordinary skill in the art, such as those described in Harlow and Lane,

Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory, 1988. For example, a polypeptide may be immobilized on a solid support and contacted with patient sera to allow binding of antibodies within the sera to the immobilized polypeptide. Unbound sera may then be removed and bound antibodies detected using, for example, ^{125}I-labeled Protein A.

As noted above, a composition may comprise a variant of a native colon tumor protein. A polypeptide "variant," as used herein, is a polypeptide that differs from a native colon tumor protein in one or more substitutions, deletions, additions and/or insertions, such that the immunogenicity of the polypeptide is not substantially diminished. In other words, the ability of a variant to react with antigen-specific antisera may be enhanced or unchanged, relative to the native protein, or may be diminished by less than 50%, and preferably less than 20%, relative to the native protein. Such variants may generally be identified by modifying one of the above polypeptide sequences and evaluating the reactivity of the modified polypeptide with antigen-specific antibodies or antisera as described herein. Preferred variants include those in which one or more portions, such as an N-terminal leader sequence or transmembrane domain, have been removed. Other preferred variants include variants in which a small portion (*e.g.,* 1-30 amino acids, preferably 5-15 amino acids) has been removed from the N- and/or C-terminal of the mature protein.

Polypeptide variants preferably exhibit at least about 70%, more preferably at least about 90% and most preferably at least about 95% identity (determined as described above) to the identified polypeptides.

Preferably, a variant contains conservative substitutions. A "conservative substitution" is one in which an amino acid is substituted for another amino acid that has similar properties, such that one skilled in the art of peptide chemistry would expect the secondary structure and hydropathic nature of the polypeptide to be substantially unchanged. Amino acid substitutions may generally be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity and/or the amphipathic nature of the residues. For example, negatively charged amino acids include aspartic acid and glutamic acid; positively charged amino acids include lysine and arginine; and amino acids with uncharged polar head groups having similar hydrophilicity values include leucine, isoleucine and valine; glycine and alanine; asparagine and glutamine; and serine, threonine, phenylalanine and tyrosine. Other groups of amino acids that may represent conservative changes include: (1) ala, pro,

gly, glu, asp, gln, asn, ser, thr; (2) cys, ser, tyr, thr; (3) val, ile, leu, met, ala, phe; (4) lys, arg, his; and (5) phe, tyr, trp, his. A variant may also, or alternatively, contain nonconservative changes. In a preferred embodiment, variant polypeptides differ from a native sequence by substitution, deletion or addition of five amino acids or fewer. Variants may also (or alternatively) be modified by, for example, the deletion or addition of amino acids that have minimal influence on the immunogenicity, secondary structure and hydropathic nature of the polypeptide.

As noted above, polypeptides may comprise a signal (or leader) sequence at the N-terminal end of the protein which co-translationally or post-translationally directs transfer of the protein. The polypeptide may also be conjugated to a linker or other sequence for ease of synthesis, purification or identification of the polypeptide (*e.g.*, poly-His), or to enhance binding of the polypeptide to a solid support. For example, a polypeptide may be conjugated to an immunoglobulin Fc region.

Polypeptides may be prepared using any of a variety of well known techniques. Recombinant polypeptides encoded by DNA sequences as described above may be readily prepared from the DNA sequences using any of a variety of expression vectors known to those of ordinary skill in the art. Expression may be achieved in any appropriate host cell that has been transformed or transfected with an expression vector containing a DNA molecule that encodes a recombinant polypeptide. Suitable host cells include prokaryotes, yeast and higher eukaryotic cells. Preferably, the host cells employed are *E. coli*, yeast or a mammalian cell line such as COS or CHO. Supernatants from suitable host/vector systems which secrete recombinant protein or polypeptide into culture media may be first concentrated using a commercially available filter. Following concentration, the concentrate may be applied to a suitable purification matrix such as an affinity matrix or an ion exchange resin. Finally, one or more reverse phase HPLC steps can be employed to further purify a recombinant polypeptide.

Portions and other variants having fewer than about 100 amino acids, and generally fewer than about 50 amino acids, may also be generated by synthetic means, using techniques well known to those of ordinary skill in the art. For example, such polypeptides may be synthesized using any of the commercially available solid-phase techniques, such as the Merrifield solid-phase synthesis method, where amino acids are sequentially added to a

growing amino acid chain. *See* Merrifield, *J. Am. Chem. Soc.* *85*:2149-2146, 1963. Equipment for automated synthesis of polypeptides is commercially available from suppliers such as Perkin Elmer/Applied BioSystems Division (Foster City, CA), and may be operated according to the manufacturer's instructions.

Within certain specific embodiments, a polypeptide may be a fusion protein that comprises multiple polypeptides as described herein, or that comprises at least one polypeptide as described herein and an unrelated sequence, such as a known tumor protein. A fusion partner may, for example, assist in providing T helper epitopes (an immunological fusion partner), preferably T helper epitopes recognized by humans, or may assist in expressing the protein (an expression enhancer) at higher yields than the native recombinant protein. Certain preferred fusion partners are both immunological and expression enhancing fusion partners. Other fusion partners may be selected so as to increase the solubility of the protein or to enable the protein to be targeted to desired intracellular compartments. Still further fusion partners include affinity tags, which facilitate purification of the protein.

Fusion proteins may generally be prepared using standard techniques, including chemical conjugation. Preferably, a fusion protein is expressed as a recombinant protein, allowing the production of increased levels, relative to a non-fused protein, in an expression system. Briefly, DNA sequences encoding the polypeptide components may be assembled separately, and ligated into an appropriate expression vector. The 3' end of the DNA sequence encoding one polypeptide component is ligated, with or without a peptide linker, to the 5' end of a DNA sequence encoding the second polypeptide component so that the reading frames of the sequences are in phase. This permits translation into a single fusion protein that retains the biological activity of both component polypeptides.

A peptide linker sequence may be employed to separate the first and the second polypeptide components by a distance sufficient to ensure that each polypeptide folds into its secondary and tertiary structures. Such a peptide linker sequence is incorporated into the fusion protein using standard techniques well known in the art. Suitable peptide linker sequences may be chosen based on the following factors: (1) their ability to adopt a flexible extended conformation; (2) their inability to adopt a secondary structure that could interact with functional epitopes on the first and second polypeptides; and (3) the lack of hydrophobic or charged residues that might react with the polypeptide functional epitopes. Preferred

peptide linker sequences contain Gly, Asn and Ser residues. Other near neutral amino acids, such as Thr and Ala may also be used in the linker sequence. Amino acid sequences which may be usefully employed as linkers include those disclosed in Maratea et al., *Gene 40*:39-46, 1985; Murphy et al., *Proc. Natl. Acad. Sci. USA 83*:8258-8262, 1986; U.S. Patent No. 4,935,233 and U.S. Patent No. 4,751,180. The linker sequence may generally be from 1 to about 50 amino acids in length. Linker sequences are not required when the first and second polypeptides have non-essential N-terminal amino acid regions that can be used to separate the functional domains and prevent steric interference.

The ligated DNA sequences are operably linked to suitable transcriptional or translational regulatory elements. The regulatory elements responsible for expression of DNA are located only 5' to the DNA sequence encoding the first polypeptides. Similarly, stop codons required to end translation and transcription termination signals are only present 3' to the DNA sequence encoding the second polypeptide.

Fusion proteins are also provided that comprise a polypeptide of the present invention together with an unrelated immunogenic protein. Preferably the immunogenic protein is capable of eliciting a recall response. Examples of such proteins include tetanus, tuberculosis and hepatitis proteins (*see*, for example, Stoute et al. *New Engl. J. Med.*, *336*:86-91, 1997).

Within preferred embodiments, an immunological fusion partner is derived from protein D, a surface protein of the gram-negative bacterium Haemophilus influenza B (WO 91/18926). Preferably, a protein D derivative comprises approximately the first third of the protein (*e.g.*, the first N-terminal 100-110 amino acids), and a protein D derivative may be lipidated. Within certain preferred embodiments, the first 109 residues of a Lipoprotein D fusion partner is included on the N-terminus to provide the polypeptide with additional exogenous T-cell epitopes and to increase the expression level in *E. coli* (thus functioning as an expression enhancer). The lipid tail ensures optimal presentation of the antigen to antigen presenting cells. Other fusion partners include the non-structural protein from influenzae virus, NS1 (hemaglutinin). Typically, the N-terminal 81 amino acids are used, although different fragments that include T-helper epitopes may be used.

In another embodiment, the immunological fusion partner is the protein known as LYTA, or a portion thereof (preferably a C-terminal portion). LYTA is derived from

Streptococcus pneumoniae, which synthesizes an N-acetyl-L-alanine amidase known as amidase LYTA (encoded by the LytA gene; *Gene 43*:265-292, 1986). LYTA is an autolysin that specifically degrades certain bonds in the peptidoglycan backbone. The C-terminal domain of the LYTA protein is responsible for the affinity to the choline or to some choline analogues such as DEAE. This property has been exploited for the development of *E. coli* C-LYTA expressing plasmids useful for expression of fusion proteins. Purification of hybrid proteins containing the C-LYTA fragment at the amino terminus has been described (*see Biotechnology 10*:795-798, 1992). Within a preferred embodiment, a repeat portion of LYTA may be incorporated into a fusion protein. A repeat portion is found in the C-terminal region starting at residue 178. A particularly preferred repeat portion incorporates residues 188-305.

In general, polypeptides (including fusion proteins) and polynucleotides as described herein are isolated. An "isolated" polypeptide or polynucleotide is one that is removed from its original environment. For example, a naturally-occurring protein is isolated if it is separated from some or all of the coexisting materials in the natural system. Preferably, such polypeptides are at least about 90% pure, more preferably at least about 95% pure and most preferably at least about 99% pure. A polynucleotide is considered to be isolated if, for example, it is cloned into a vector that is not a part of the natural environment.

BINDING AGENTS

The present invention further provides agents, such as antibodies and antigen-binding fragments thereof, that specifically bind to a colon tumor protein. As used herein, an antibody, or antigen-binding fragment thereof, is said to "specifically bind" to a colon tumor protein if it reacts at a detectable level (within, for example, an ELISA) with a colon tumor protein, and does not react detectably with unrelated proteins under similar conditions. As used herein, "binding" refers to a noncovalent association between two separate molecules such that a complex is formed. The ability to bind may be evaluated by, for example, determining a binding constant for the formation of the complex. The binding constant is the value obtained when the concentration of the complex is divided by the product of the component concentrations. In general, two compounds are said to "bind," in the context of the present invention, when the binding constant for complex formation exceeds about 10^3 L/mol. The binding constant may be determined using methods well known in the art.

Binding agents may be further capable of differentiating between patients with and without a cancer, such as colon cancer, using the representative assays provided herein. In other words, antibodies or other binding agents that bind to a colon tumor protein will generate a signal indicating the presence of a cancer in at least about 20% of patients with the disease, and will generate a negative signal indicating the absence of the disease in at least about 90% of individuals without the cancer. To determine whether a binding agent satisfies this requirement, biological samples (*e.g.,* blood, sera, sputum, urine and/or tumor biopsies) from patients with and without a cancer (as determined using standard clinical tests) may be assayed as described herein for the presence of polypeptides that bind to the binding agent. It will be apparent that a statistically significant number of samples with and without the disease should be assayed. Each binding agent should satisfy the above criteria; however, those of ordinary skill in the art will recognize that binding agents may be used in combination to improve sensitivity.

Any agent that satisfies the above requirements may be a binding agent. For example, a binding agent may be a ribosome, with or without a peptide component, an RNA molecule or a polypeptide. In a preferred embodiment, a binding agent is an antibody or an antigen-binding fragment thereof. Antibodies may be prepared by any of a variety of techniques known to those of ordinary skill in the art. *See, e.g.,* Harlow and Lane, *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory, 1988. In general, antibodies can be produced by cell culture techniques, including the generation of monoclonal antibodies as described herein, or via transfection of antibody genes into suitable bacterial or mammalian cell hosts, in order to allow for the production of recombinant antibodies. In one technique, an immunogen comprising the polypeptide is initially injected into any of a wide variety of mammals (*e.g.,* mice, rats, rabbits, sheep or goats). In this step, the polypeptides of this invention may serve as the immunogen without modification. Alternatively, particularly for relatively short polypeptides, a superior immune response may be elicited if the polypeptide is joined to a carrier protein, such as bovine serum albumin or keyhole limpet hemocyanin. The immunogen is injected into the animal host, preferably according to a predetermined schedule incorporating one or more booster immunizations, and the animals are bled periodically. Polyclonal antibodies specific for the polypeptide may then

be purified from such antisera by, for example, affinity chromatography using the polypeptide coupled to a suitable solid support.

Monoclonal antibodies specific for an antigenic polypeptide of interest may be prepared, for example, using the technique of Kohler and Milstein, *Eur. J. Immunol. 6*:511-519, 1976, and improvements thereto. Briefly, these methods involve the preparation of immortal cell lines capable of producing antibodies having the desired specificity (*i.e.*, reactivity with the polypeptide of interest). Such cell lines may be produced, for example, from spleen cells obtained from an animal immunized as described above. The spleen cells are then immortalized by, for example, fusion with a myeloma cell fusion partner, preferably one that is syngeneic with the immunized animal. A variety of fusion techniques may be employed. For example, the spleen cells and myeloma cells may be combined with a nonionic detergent for a few minutes and then plated at low density on a selective medium that supports the growth of hybrid cells, but not myeloma cells. A preferred selection technique uses HAT (hypoxanthine, aminopterin, thymidine) selection. After a sufficient time, usually about 1 to 2 weeks, colonies of hybrids are observed. Single colonies are selected and their culture supernatants tested for binding activity against the polypeptide. Hybridomas having high reactivity and specificity are preferred.

Monoclonal antibodies may be isolated from the supernatants of growing hybridoma colonies. In addition, various techniques may be employed to enhance the yield, such as injection of the hybridoma cell line into the peritoneal cavity of a suitable vertebrate host, such as a mouse. Monoclonal antibodies may then be harvested from the ascites fluid or the blood. Contaminants may be removed from the antibodies by conventional techniques, such as chromatography, gel filtration, precipitation, and extraction. The polypeptides of this invention may be used in the purification process in, for example, an affinity chromatography step.

Within certain embodiments, the use of antigen-binding fragments of antibodies may be preferred. Such fragments include Fab fragments, which may be prepared using standard techniques. Briefly, immunoglobulins may be purified from rabbit serum by affinity chromatography on Protein A bead columns (Harlow and Lane, *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory, 1988) and digested by papain to yield

Fab and Fc fragments. The Fab and Fc fragments may be separated by affinity chromatography on protein A bead columns.

Monoclonal antibodies of the present invention may be coupled to one or more therapeutic agents. Suitable agents in this regard include radionuclides, differentiation inducers, drugs, toxins, and derivatives thereof. Preferred radionuclides include ^{90}Y, ^{123}I, ^{125}I, ^{131}I, ^{186}Re, ^{188}Re, ^{211}At, and ^{212}Bi. Preferred drugs include methotrexate, and pyrimidine and purine analogs. Preferred differentiation inducers include phorbol esters and butyric acid. Preferred toxins include ricin, abrin, diptheria toxin, cholera toxin, gelonin, Pseudomonas exotoxin, Shigella toxin, and pokeweed antiviral protein.

A therapeutic agent may be coupled (*e.g.*, covalently bonded) to a suitable monoclonal antibody either directly or indirectly (*e.g.*, via a linker group). A direct reaction between an agent and an antibody is possible when each possesses a substituent capable of reacting with the other. For example, a nucleophilic group, such as an amino or sulfhydryl group, on one may be capable of reacting with a carbonyl-containing group, such as an anhydride or an acid halide, or with an alkyl group containing a good leaving group (*e.g.*, a halide) on the other.

Alternatively, it may be desirable to couple a therapeutic agent and an antibody via a linker group. A linker group can function as a spacer to distance an antibody from an agent in order to avoid interference with binding capabilities. A linker group can also serve to increase the chemical reactivity of a substituent on an agent or an antibody, and thus increase the coupling efficiency. An increase in chemical reactivity may also facilitate the use of agents, or functional groups on agents, which otherwise would not be possible.

It will be evident to those skilled in the art that a variety of bifunctional or polyfunctional reagents, both homo- and hetero-functional (such as those described in the catalog of the Pierce Chemical Co., Rockford, IL), may be employed as the linker group. Coupling may be effected, for example, through amino groups, carboxyl groups, sulfhydryl groups or oxidized carbohydrate residues. There are numerous references describing such methodology, *e.g.*, U.S. Patent No. 4,671,958, to Rodwell et al.

Where a therapeutic agent is more potent when free from the antibody portion of the immunoconjugates of the present invention, it may be desirable to use a linker group which is cleavable during or upon internalization into a cell. A number of different cleavable

linker groups have been described. The mechanisms for the intracellular release of an agent from these linker groups include cleavage by reduction of a disulfide bond (*e.g.*, U.S. Patent No. 4,489,710, to Spitler), by irradiation of a photolabile bond (*e.g.*, U.S. Patent No. 4,625,014, to Senter et al.), by hydrolysis of derivatized amino acid side chains (*e.g.*, U.S. Patent No. 4,638,045, to Kohn et al.), by serum complement-mediated hydrolysis (*e.g.*, U.S. Patent No. 4,671,958, to Rodwell et al.), and acid-catalyzed hydrolysis (*e.g.*, U.S. Patent No. 4,569,789, to Blattler et al.).

It may be desirable to couple more than one agent to an antibody. In one embodiment, multiple molecules of an agent are coupled to one antibody molecule. In another embodiment, more than one type of agent may be coupled to one antibody. Regardless of the particular embodiment, immunoconjugates with more than one agent may be prepared in a variety of ways. For example, more than one agent may be coupled directly to an antibody molecule, or linkers which provide multiple sites for attachment can be used. Alternatively, a carrier can be used.

A carrier may bear the agents in a variety of ways, including covalent bonding either directly or via a linker group. Suitable carriers include proteins such as albumins (*e.g.*, U.S. Patent No. 4,507,234, to Kato et al.), peptides and polysaccharides such as aminodextran (*e.g.*, U.S. Patent No. 4,699,784, to Shih et al.). A carrier may also bear an agent by noncovalent bonding or by encapsulation, such as within a liposome vesicle (*e.g.*, U.S. Patent Nos. 4,429,008 and 4,873,088). Carriers specific for radionuclide agents include radiohalogenated small molecules and chelating compounds. For example, U.S. Patent No. 4,735,792 discloses representative radiohalogenated small molecules and their synthesis. A radionuclide chelate may be formed from chelating compounds that include those containing nitrogen and sulfur atoms as the donor atoms for binding the metal, or metal oxide, radionuclide. For example, U.S. Patent No. 4,673,562, to Davison et al. discloses representative chelating compounds and their synthesis.

A variety of routes of administration for the antibodies and immunoconjugates may be used. Typically, administration will be intravenous, intramuscular, subcutaneous or in the bed of a resected tumor. It will be evident that the precise dose of the antibody/immunoconjugate will vary depending upon the antibody used, the antigen density on the tumor, and the rate of clearance of the antibody.

T CELLS

Immunotherapeutic compositions may also, or alternatively, comprise T cells specific for a colon tumor protein. Such cells may generally be prepared *in vitro* or *ex vivo*, using standard procedures. For example, T cells may be isolated from bone marrow, peripheral blood, or a fraction of bone marrow or peripheral blood of a patient, using a commercially available cell separation system, such as the CEPRATE™ system, available from CellPro Inc., Bothell WA (see also U.S. Patent No. 5,240,856; U.S. Patent No. 5,215,926; WO 89/06280; WO 91/16116 and WO 92/07243). Alternatively, T cells may be derived from related or unrelated humans, non-human mammals, cell lines or cultures.

T cells may be stimulated with a colon tumor polypeptide, polynucleotide encoding a colon tumor polypeptide and/or an antigen presenting cell (APC) that expresses such a polypeptide. Such stimulation is performed under conditions and for a time sufficient to permit the generation of T cells that are specific for the polypeptide. Preferably, a colon tumor polypeptide or polynucleotide is present within a delivery vehicle, such as a microsphere, to facilitate the generation of specific T cells.

T cells are considered to be specific for a colon tumor polypeptide if the T cells kill target cells coated with the polypeptide or expressing a gene encoding the polypeptide. T cell specificity may be evaluated using any of a variety of standard techniques. For example, within a chromium release assay or proliferation assay, a stimulation index of more than two fold increase in lysis and/or proliferation, compared to negative controls, indicates T cell specificity. Such assays may be performed, for example, as described in Chen et al., *Cancer Res. 54*:1065-1070, 1994. Alternatively, detection of the proliferation of T cells may be accomplished by a variety of known techniques. For example, T cell proliferation can be detected by measuring an increased rate of DNA synthesis (*e.g.*, by pulse-labeling cultures of T cells with tritiated thymidine and measuring the amount of tritiated thymidine incorporated into DNA). Contact with a colon tumor polypeptide (100 ng/ml - 100 μg/ml, preferably 200 ng/ml - 25 μg/ml) for 3 - 7 days should result in at least a two fold increase in proliferation of the T cells. Contact as described above for 2-3 hours should result in activation of the T cells, as measured using standard cytokine assays in which a two fold increase in the level of cytokine release (*e.g.*, TNF or IFN-γ) is indicative of T cell

activation (*see* Coligan et al., Current Protocols in Immunology, vol. 1, Wiley Interscience (Greene 1998)). T cells that have been activated in response to a colon tumor polypeptide, polynucleotide or polypeptide-expressing APC may be $CD4^+$ and/or $CD8^-$. Colon tumor protein-specific T cells may be expanded using standard techniques. Within preferred embodiments, the T cells are derived from either a patient or a related, or unrelated, donor and are administered to the patient following stimulation and expansion.

For therapeutic purposes, $CD4^+$ or $CD8^+$ T cells that proliferate in response to a colon tumor polypeptide, polynucleotide or APC can be expanded in number either *in vitro* or *in vivo*. Proliferation of such T cells *in vitro* may be accomplished in a variety of ways. For example, the T cells can be re-exposed to a colon tumor polypeptide, or a short peptide corresponding to an immunogenic portion of such a polypeptide, with or without the addition of T cell growth factors, such as interleukin-2, and/or stimulator cells that synthesize a colon tumor polypeptide. Alternatively, one or more T cells that proliferate in the presence of a colon tumor protein can be expanded in number by cloning. Methods for cloning cells are well known in the art, and include limiting dilution.

PHARMACEUTICAL COMPOSITIONS AND VACCINES

Within certain aspects, polypeptides, polynucleotides, T cells and/or binding agents disclosed herein may be incorporated into pharmaceutical compositions or immunogenic compositions (*i.e.*, vaccines). Pharmaceutical compositions comprise one or more such compounds and a physiologically acceptable carrier. Vaccines may comprise one or more such compounds and a non-specific immune response enhancer. A non-specific immune response enhancer may be any substance that enhances an immune response to an exogenous antigen. Examples of non-specific immune response enhancers include adjuvants, biodegradable microspheres (*e.g.*, polylactic galactide) and liposomes (into which the compound is incorporated; *see e.g.*, Fullerton, U.S. Patent No. 4,235,877). Vaccine preparation is generally described in, for example, M.F. Powell and M.J. Newman, eds., "Vaccine Design (the subunit and adjuvant approach)," Plenum Press (NY, 1995). Pharmaceutical compositions and vaccines within the scope of the present invention may also contain other compounds, which may be biologically active or inactive. For example, one or

more immunogenic portions of other tumor antigens may be present, either incorporated into a fusion polypeptide or as a separate compound, within the composition or vaccine.

A pharmaceutical composition or vaccine may contain DNA encoding one or more of the polypeptides as described above, such that the polypeptide is generated *in situ*. As noted above, the DNA may be present within any of a variety of delivery systems known to those of ordinary skill in the art, including nucleic acid expression systems, bacteria and viral expression systems. Numerous gene delivery techniques are well known in the art, such as those described by Rolland, *Crit. Rev. Therap. Drug Carrier Systems 15*:143-198, 1998, and references cited therein. Appropriate nucleic acid expression systems contain the necessary DNA sequences for expression in the patient (such as a suitable promoter and terminating signal). Bacterial delivery systems involve the administration of a bacterium (such as *Bacillus-Calmette-Guerrin*) that expresses an immunogenic portion of the polypeptide on its cell surface or secretes such an epitope. In a preferred embodiment, the DNA may be introduced using a viral expression system (*e.g.,* vaccinia or other pox virus, retrovirus, or adenovirus), which may involve the use of a non-pathogenic (defective), replication competent virus. Suitable systems are disclosed, for example, in Fisher-Hoch et al., *Proc. Natl. Acad. Sci. USA 86*:317-321, 1989; Flexner et al., *Ann. N.Y. Acad. Sci. 569*:86-103, 1989; Flexner et al., *Vaccine 8*:17-21, 1990; U.S. Patent Nos. 4,603,112, 4,769,330, and 5,017,487; WO 89/01973; U.S. Patent No. 4,777,127; GB 2,200,651; EP 0,345,242; WO 91/02805; Berkner, *Biotechniques 6*:616-627, 1988; Rosenfeld et al., *Science 252*:431-434, 1991; Kolls et al., *Proc. Natl. Acad. Sci. USA 91*:215-219, 1994; Kass-Eisler et al., *Proc. Natl. Acad. Sci. USA 90*:11498-11502, 1993; Guzman et al., *Circulation 88*:2838-2848, 1993; and Guzman et al., *Cir. Res. 73*:1202-1207, 1993. Techniques for incorporating DNA into such expression systems are well known to those of ordinary skill in the art. The DNA may also be "naked," as described, for example, in Ulmer et al., *Science 259*:1745-1749, 1993 and reviewed by Cohen, *Science 259*:1691-1692, 1993. The uptake of naked DNA may be increased by coating the DNA onto biodegradable beads, which are efficiently transported into the cells.

While any suitable carrier known to those of ordinary skill in the art may be employed in the pharmaceutical compositions of this invention, the type of carrier will vary depending on the mode of administration. Compositions of the present invention may be

formulated for any appropriate manner of administration, including for example, topical, oral, nasal, intravenous, intracranial, intraperitoneal, subcutaneous or intramuscular administration. For parenteral administration, such as subcutaneous injection, the carrier preferably comprises water, saline, alcohol, a fat, a wax or a buffer. For oral administration, any of the above carriers or a solid carrier, such as mannitol, lactose, starch, magnesium stearate, sodium saccharine, talcum, cellulose, glucose, sucrose, and magnesium carbonate, may be employed. Biodegradable microspheres (*e.g.,* polylactate polyglycolate) may also be employed as carriers for the pharmaceutical compositions of this invention. Suitable biodegradable microspheres are disclosed, for example, in U.S. Patent Nos. 4,897,268 and 5,075,109.

Such compositions may also comprise buffers (*e.g.,* neutral buffered saline or phosphate buffered saline), carbohydrates (*e.g.,* glucose, mannose, sucrose or dextrans), mannitol, proteins, polypeptides or amino acids such as glycine, antioxidants, chelating agents such as EDTA or glutathione, adjuvants (*e.g.,* aluminum hydroxide) and/or preservatives. Alternatively, compositions of the present invention may be formulated as a lyophilizate. Compounds may also be encapsulated within liposomes using well known technology.

Any of a variety of non-specific immune response enhancers may be employed in the vaccines of this invention. For example, an adjuvant may be included. Most adjuvants contain a substance designed to protect the antigen from rapid catabolism, such as aluminum hydroxide or mineral oil, and a stimulator of immune responses, such as lipid A, *Bortadella pertussis* or *Mycobacterium tuberculosis* derived proteins. Suitable adjuvants are commercially available as, for example, Freund's Incomplete Adjuvant and Complete Adjuvant (Difco Laboratories, Detroit, MI); Merck Adjuvant 65 (Merck and Company, Inc., Rahway, NJ); aluminum salts such as aluminum hydroxide gel (alum) or aluminum phosphate; salts of calcium, iron or zinc; an insoluble suspension of acylated tyrosine; acylated sugars; cationically or anionically derivatized polysaccharides; polyphosphazenes; biodegradable microspheres; monophosphoryl lipid A and quil A. Cytokines, such as GM-CSF or interleukin-2, -7, or -12, may also be used as adjuvants.

Within the vaccines provided herein, the adjuvant composition is preferably designed to induce an immune response predominantly of the Th1 type. High levels of Th1-

type cytokines (*e.g.*, IFN-γ, IL-2 and IL-12) tend to favor the induction of cell mediated immune responses to an administered antigen. In contrast, high levels of Th2-type cytokines (*e.g.*, IL-4, IL-5, IL-6, IL-10 and TNF-β) tend to favor the induction of humoral immune responses. Following application of a vaccine as provided herein, a patient will support an immune response that includes Th1- and Th2-type responses. Within a preferred embodiment, in which a response is predominantly Th1-type, the level of Th1-type cytokines will increase to a greater extent than the level of Th2-type cytokines. The levels of these cytokines may be readily assessed using standard assays. For a review of the families of cytokines, see Mosmann and Coffman, *Ann. Rev. Immunol.* 7:145-173, 1989.

Preferred adjuvants for use in eliciting a predominantly Th1-type response include, for example, a combination of monophosphoryl lipid A, preferably 3-de-O-acylated monophosphoryl lipid A (3D-MPL), together with an aluminum salt. MPL adjuvants are available from Ribi ImmunoChem Research Inc. (Hamillton, MT) (*see* US Patent Nos. 4,436,727; 4,877,611; 4,866,034 and 4,912,094). CpG-containing oligonucleotides (in which the CpG dinucleotide is unmethylated) also induce a predominantly Th1 response. Such oligonucleotides are well known and are described, for example, in WO 96/02555. Another preferred adjuvant is a saponin, preferably QS21, which may be used alone or in combination with other adjuvants. For example, an enhanced system involves the combination of a monophosphoryl lipid A and saponin derivative, such as the combination of QS21 and 3D-MPL as described in WO 94/00153, or a less reactogenic composition where the QS21 is quenched with cholesterol, as described in WO 96/33739. Other preferred formulations comprises an oil-in-water emulsion and tocopherol. A particularly potent adjuvant formulation involving QS21, 3D-MPL and tocopherol in an oil-in-water emulsion is described in WO 95/17210. Any vaccine provided herein may be prepared using well known methods that result in a combination of antigen, immune response enhancer and a suitable carrier or excipient.

The compositions described herein may be administered as part of a sustained release formulation (*i.e.*, a formulation such as a capsule or sponge that effects a slow release of compound following administration). Such formulations may generally be prepared using well known technology and administered by, for example, oral, rectal or subcutaneous implantation, or by implantation at the desired target site. Sustained-release formulations

may contain a polypeptide, polynucleotide or antibody dispersed in a carrier matrix and/or contained within a reservoir surrounded by a rate controlling membrane. Carriers for use within such formulations are biocompatible, and may also be biodegradable; preferably the formulation provides a relatively constant level of active component release. The amount of active compound contained within a sustained release formulation depends upon the site of implantation, the rate and expected duration of release and the nature of the condition to be treated or prevented.

Any of a variety of delivery vehicles may be employed within pharmaceutical compositions and vaccines to facilitate production of an antigen-specific immune response that targets tumor cells. Delivery vehicles include antigen presenting cells (APCs), such as dendritic cells, macrophages, B cells, monocytes and other cells that may be engineered to be efficient APCs. Such cells may, but need not, be genetically modified to increase the capacity for presenting the antigen, to improve activation and/or maintenance of the T cell response, to have anti-tumor effects *per se* and/or to be immunologically compatible with the receiver (*i.e.,* matched HLA haplotype). APCs may generally be isolated from any of a variety of biological fluids and organs, including tumor and peritumoral tissues, and may be autologous, allogeneic, syngeneic or xenogeneic cells.

Certain preferred embodiments of the present invention use dendritic cells or progenitors thereof as antigen-presenting cells. Dendritic cells are highly potent APCs (Banchereau and Steinman, *Nature 392*:245-251, 1998) and have been shown to be effective as a physiological adjuvant for eliciting prophylactic or therapeutic antitumor immunity (*see* Timmerman and Levy, *Ann. Rev. Med. 50*:507-529, 1999). In general, dendritic cells may be identified based on their typical shape (stellate *in situ*, with marked cytoplasmic processes (dendrites) visible *in vitro*) and based on the lack of differentiation markers of B cells (CD19 and CD20), T cells (CD3), monocytes (CD14) and natural killer cells (CD56), as determined using standard assays. Dendritic cells may, of course, be engineered to express specific cell-surface receptors or ligands that are not commonly found on dendritic cells *in vivo* or *ex vivo*, and such modified dendritic cells are contemplated by the present invention. As an alternative to dendritic cells, secreted vesicles antigen-loaded dendritic cells (called exosomes) may be used within a vaccine (*see* Zitvogel et al., *Nature Med. 4*:594-600, 1998).

Dendritic cells and progenitors may be obtained from peripheral blood, bone marrow, tumor-infiltrating cells, peritumoral tissues-infiltrating cells, lymph nodes, spleen, skin, umbilical cord blood or any other suitable tissue or fluid. For example, dendritic cells may be differentiated *ex vivo* by adding a combination of cytokines such as GM-CSF, IL-4, IL-13 and/or TNFα to cultures of monocytes harvested from peripheral blood. Alternatively, CD34 positive cells harvested from peripheral blood, umbilical cord blood or bone marrow may be differentiated into dendritic cells by adding to the culture medium combinations of GM-CSF, IL-3, TNFα, CD40 ligand, LPS, flt3 ligand and/or other compound(s) that induce maturation and proliferation of dendritic cells.

Dendritic cells are conveniently categorized as "immature" and "mature" cells, which allows a simple way to discriminate between two well characterized phenotypes. However, this nomenclature should not be construed to exclude all possible intermediate stages of differentiation. Immature dendritic cells are characterized as APC with a high capacity for antigen uptake and processing, which correlates with the high expression of Fcγ receptor, mannose receptor and DEC-205 marker. The mature phenotype is typically characterized by a lower expression of these markers, but a high expression of cell surface molecules responsible for T cell activation such as class I and class II MHC, adhesion molecules (*e.g.,* CD54 and CD11) and costimulatory molecules (*e.g.,* CD40, CD80 and CD86).

APCs may generally be transfected with a polynucleotide encoding a colon tumor protein (or portion or other variant thereof) such that the colon tumor polypeptide, or an immunogenic portion thereof, is expressed on the cell surface. Such transfection may take place *ex vivo*, and a composition or vaccine comprising such transfected cells may then be used for therapeutic purposes, as described herein. Alternatively, a gene delivery vehicle that targets a dendritic or other antigen presenting cell may be administered to a patient, resulting in transfection that occurs *in vivo*. *In vivo* and *ex vivo* transfection of dendritic cells, for example, may generally be performed using any methods known in the art, such as those described in WO 97/24447, or the gene gun approach described by Mahvi et al., *Immunology and cell Biology* 75:456-460, 1997. Antigen loading of dendritic cells may be achieved by incubating dendritic cells or progenitor cells with the colon tumor polypeptide, DNA (naked or within a plasmid vector) or RNA; or with antigen-expressing recombinant bacterium or

viruses (*e.g.*, vaccinia, fowlpox, adenovirus or lentivirus vectors). Prior to loading, the polypeptide may be covalently conjugated to an immunological partner that provides T cell help (*e.g.*, a carrier molecule). Alternatively, a dendritic cell may be pulsed with a non-conjugated immunological partner, separately or in the presence of the polypeptide.

CANCER THERAPY

In further aspects of the present invention, the compositions described herein may be used for immunotherapy of cancer, such as colon cancer. Within such methods, pharmaceutical compositions and vaccines are typically administered to a patient. As used herein, a "patient" refers to any warm-blooded animal, preferably a human. A patient may or may not be afflicted with cancer. Accordingly, the above pharmaceutical compositions and vaccines may be used to prevent the development of a cancer or to treat a patient afflicted with a cancer. A cancer may be diagnosed using criteria generally accepted in the art, including the presence of a malignant tumor. Pharmaceutical compositions and vaccines may be administered either prior to or following surgical removal of primary tumors and/or treatment such as administration of radiotherapy or conventional chemotherapeutic drugs.

Within certain embodiments, immunotherapy may be active immunotherapy, in which treatment relies on the *in vivo* stimulation of the endogenous host immune system to react against tumors with the administration of immune response-modifying agents (such as polypeptides and polynucleotides disclosed herein).

Within other embodiments, immunotherapy may be passive immunotherapy, in which treatment involves the delivery of agents with established tumor-immune reactivity (such as effector cells or antibodies) that can directly or indirectly mediate antitumor effects and does not necessarily depend on an intact host immune system. Examples of effector cells include T cells as discussed above, T lymphocytes (such as $CD8^+$ cytotoxic T lymphocytes and $CD4^+$ T-helper tumor-infiltrating lymphocytes), killer cells (such as Natural Killer cells and lymphokine-activated killer cells), B cells and antigen-presenting cells (such as dendritic cells and macrophages) expressing a polypeptide provided herein. T cell receptors and antibody receptors specific for the polypeptides recited herein may be cloned, expressed and transferred into other vectors or effector cells for adoptive immunotherapy. The polypeptides

provided herein may also be used to generate antibodies or anti-idiotypic antibodies (as described above and in U.S. Patent No. 4,918,164) for passive immunotherapy.

Effector cells may generally be obtained in sufficient quantities for adoptive immunotherapy by growth *in vitro*, as described herein. Culture conditions for expanding single antigen-specific effector cells to several billion in number with retention of antigen recognition *in vivo* are well known in the art. Such *in vitro* culture conditions typically use intermittent stimulation with antigen, often in the presence of cytokines (such as IL-2) and non-dividing feeder cells. As noted above, immunoreactive polypeptides as provided herein may be used to rapidly expand antigen-specific T cell cultures in order to generate a sufficient number of cells for immunotherapy. In particular, antigen-presenting cells, such as dendritic, macrophage, monocyte, fibroblast or B cells, may be pulsed with immunoreactive polypeptides or transfected with one or more polynucleotides using standard techniques well known in the art. For example, antigen-presenting cells can be transfected with a polynucleotide having a promoter appropriate for increasing expression in a recombinant virus or other expression system. Cultured effector cells for use in therapy must be able to grow and distribute widely, and to survive long term *in vivo*. Studies have shown that cultured effector cells can be induced to grow in vivo and to survive long term in substantial numbers by repeated stimulation with antigen supplemented with IL-2 (*see*, for example, Cheever et al., *Immunological Reviews 157*:177, 1997).

Alternatively, a vector expressing a polypeptide recited herein may be introduced into antigen presenting cells taken from a patient and clonally propagated *ex vivo* for transplant back into the same patient. Transfected cells may be reintroduced into the patient using any means known in the art, preferably in sterile form by intravenous, intracavitary, intraperitoneal or intratumor administration.

Routes and frequency of administration of the therapeutic compositions disclosed herein, as well as dosage, will vary from individual to individual, and may be readily established using standard techniques. In general, the pharmaceutical compositions and vaccines may be administered by injection (*e.g.*, intracutaneous, intramuscular, intravenous or subcutaneous), intranasally (*e.g.*, by aspiration) or orally. Preferably, between 1 and 10 doses may be administered over a 52 week period. Preferably, 6 doses are administered, at intervals of 1 month, and booster vaccinations may be given periodically

thereafter. Alternate protocols may be appropriate for individual patients. A suitable dose is an amount of a compound that, when administered as described above, is capable of promoting an anti-tumor immune response, and is at least 10-50% above the basal (*i.e.*, untreated) level. Such response can be monitored by measuring the anti-tumor antibodies in a patient or by vaccine-dependent generation of cytolytic effector cells capable of killing the patient's tumor cells *in vitro*. Such vaccines should also be capable of causing an immune response that leads to an improved clinical outcome (*e.g.*, more frequent remissions, complete or partial or longer disease-free survival) in vaccinated patients as compared to non-vaccinated patients. In general, for pharmaceutical compositions and vaccines comprising one or more polypeptides, the amount of each polypeptide present in a dose ranges from about 100 μg to 5 mg per kg of host. Suitable dose sizes will vary with the size of the patient, but will typically range from about 0.1 mL to about 5 mL.

In general, an appropriate dosage and treatment regimen provides the active compound(s) in an amount sufficient to provide therapeutic and/or prophylactic benefit. Such a response can be monitored by establishing an improved clinical outcome (*e.g.*, more frequent remissions, complete or partial, or longer disease-free survival) in treated patients as compared to non-treated patients. Increases in preexisting immune responses to a colon tumor protein generally correlate with an improved clinical outcome. Such immune responses may generally be evaluated using standard proliferation, cytotoxicity or cytokine assays, which may be performed using samples obtained from a patient before and after treatment.

METHODS FOR DETECTING CANCER

In general, a cancer may be detected in a patient based on the presence of one or more colon tumor proteins and/or polynucleotides encoding such proteins in a biological sample (for example, blood, sera, sputum, urine and/or tumor biopsies) obtained from the patient. In other words, such proteins may be used as markers to indicate the presence or absence of a cancer such as colon cancer. In addition, such proteins may be useful for the detection of other cancers. The binding agents provided herein generally permit detection of the level of antigen that binds to the agent in the biological sample. Polynucleotide primers and probes may be used to detect the level of mRNA encoding a tumor protein, which is also

indicative of the presence or absence of a cancer. In general, a colon tumor sequence should be present at a level that is at least three fold higher in tumor tissue than in normal tissue

There are a variety of assay formats known to those of ordinary skill in the art for using a binding agent to detect polypeptide markers in a sample. *See*, *e.g.*, Harlow and Lane, *Antibodies: A Laboratory Manual,* Cold Spring Harbor Laboratory, 1988. In general, the presence or absence of a cancer in a patient may be determined by (a) contacting a biological sample obtained from a patient with a binding agent; (b) detecting in the sample a level of polypeptide that binds to the binding agent; and (c) comparing the level of polypeptide with a predetermined cut-off value.

In a preferred embodiment, the assay involves the use of binding agent immobilized on a solid support to bind to and remove the polypeptide from the remainder of the sample. The bound polypeptide may then be detected using a detection reagent that contains a reporter group and specifically binds to the binding agent/polypeptide complex. Such detection reagents may comprise, for example, a binding agent that specifically binds to the polypeptide or an antibody or other agent that specifically binds to the binding agent, such as an anti-immunoglobulin, protein G. protein A or a lectin. Alternatively, a competitive assay may be utilized, in which a polypeptide is labeled with a reporter group and allowed to bind to the immobilized binding agent after incubation of the binding agent with the sample. The extent to which components of the sample inhibit the binding of the labeled polypeptide to the binding agent is indicative of the reactivity of the sample with the immobilized binding agent. Suitable polypeptides for use within such assays include full length colon tumor proteins and portions thereof to which the binding agent binds, as described above.

The solid support may be any material known to those of ordinary skill in the art to which the tumor protein may be attached. For example, the solid support may be a test well in a microtiter plate or a nitrocellulose or other suitable membrane. Alternatively, the support may be a bead or disc, such as glass, fiberglass, latex or a plastic material such as polystyrene or polyvinylchloride. The support may also be a magnetic particle or a fiber optic sensor, such as those disclosed, for example, in U.S. Patent No. 5,359,681. The binding agent may be immobilized on the solid support using a variety of techniques known to those of skill in the art, which are amply described in the patent and scientific literature. In the context of the present invention, the term "immobilization" refers to both noncovalent

association, such as adsorption, and covalent attachment (which may be a direct linkage between the agent and functional groups on the support or may be a linkage by way of a cross-linking agent). Immobilization by adsorption to a well in a microtiter plate or to a membrane is preferred. In such cases, adsorption may be achieved by contacting the binding agent, in a suitable buffer, with the solid support for a suitable amount of time. The contact time varies with temperature, but is typically between about 1 hour and about 1 day. In general, contacting a well of a plastic microtiter plate (such as polystyrene or polyvinylchloride) with an amount of binding agent ranging from about 10 ng to about 10 μg, and preferably about 100 ng to about 1 μg, is sufficient to immobilize an adequate amount of binding agent.

Covalent attachment of binding agent to a solid support may generally be achieved by first reacting the support with a bifunctional reagent that will react with both the support and a functional group, such as a hydroxyl or amino group, on the binding agent. For example, the binding agent may be covalently attached to supports having an appropriate polymer coating using benzoquinone or by condensation of an aldehyde group on the support with an amine and an active hydrogen on the binding partner (*see*, *e.g.*, Pierce Immunotechnology Catalog and Handbook, 1991, at A12-A13).

In certain embodiments, the assay is a two-antibody sandwich assay. This assay may be performed by first contacting an antibody that has been immobilized on a solid support, commonly the well of a microtiter plate, with the sample, such that polypeptides within the sample are allowed to bind to the immobilized antibody. Unbound sample is then removed from the immobilized polypeptide-antibody complexes and a detection reagent (preferably a second antibody capable of binding to a different site on the polypeptide) containing a reporter group is added. The amount of detection reagent that remains bound to the solid support is then determined using a method appropriate for the specific reporter group.

More specifically, once the antibody is immobilized on the support as described above, the remaining protein binding sites on the support are typically blocked. Any suitable blocking agent known to those of ordinary skill in the art, such as bovine serum albumin or Tween 20™ (Sigma Chemical Co., St. Louis, MO). The immobilized antibody is then incubated with the sample, and polypeptide is allowed to bind to the antibody. The

sample may be diluted with a suitable diluent, such as phosphate-buffered saline (PBS) prior to incubation. In general, an appropriate contact time (*i.e.*, incubation time) is a period of time that is sufficient to detect the presence of polypeptide within a sample obtained from an individual with colon cancer. Preferably, the contact time is sufficient to achieve a level of binding that is at least about 95% of that achieved at equilibrium between bound and unbound polypeptide. Those of ordinary skill in the art will recognize that the time necessary to achieve equilibrium may be readily determined by assaying the level of binding that occurs over a period of time. At room temperature, an incubation time of about 30 minutes is generally sufficient.

Unbound sample may then be removed by washing the solid support with an appropriate buffer, such as PBS containing 0.1% Tween 20™. The second antibody, which contains a reporter group, may then be added to the solid support. Preferred reporter groups include those groups recited above.

The detection reagent is then incubated with the immobilized antibody-polypeptide complex for an amount of time sufficient to detect the bound polypeptide. An appropriate amount of time may generally be determined by assaying the level of binding that occurs over a period of time. Unbound detection reagent is then removed and bound detection reagent is detected using the reporter group. The method employed for detecting the reporter group depends upon the nature of the reporter group. For radioactive groups, scintillation counting or autoradiographic methods are generally appropriate. Spectroscopic methods may be used to detect dyes, luminescent groups and fluorescent groups. Biotin may be detected using avidin, coupled to a different reporter group (commonly a radioactive or fluorescent group or an enzyme). Enzyme reporter groups may generally be detected by the addition of substrate (generally for a specific period of time), followed by spectroscopic or other analysis of the reaction products.

To determine the presence or absence of a cancer, such as colon cancer, the signal detected from the reporter group that remains bound to the solid support is generally compared to a signal that corresponds to a predetermined cut-off value. In one preferred embodiment, the cut-off value for the detection of a cancer is the average mean signal obtained when the immobilized antibody is incubated with samples from patients without the cancer. In general, a sample generating a signal that is three standard deviations above the

predetermined cut-off value is considered positive for the cancer. In an alternate preferred embodiment, the cut-off value is determined using a Receiver Operator Curve, according to the method of Sackett et al., *Clinical Epidemiology: A Basic Science for Clinical Medicine*, Little Brown and Co., 1985, p. 106-7. Briefly, in this embodiment, the cut-off value may be determined from a plot of pairs of true positive rates (*i.e.*, sensitivity) and false positive rates (100%-specificity) that correspond to each possible cut-off value for the diagnostic test result. The cut-off value on the plot that is the closest to the upper left-hand corner (*i.e.*, the value that encloses the largest area) is the most accurate cut-off value, and a sample generating a signal that is higher than the cut-off value determined by this method may be considered positive. Alternatively, the cut-off value may be shifted to the left along the plot, to minimize the false positive rate, or to the right, to minimize the false negative rate. In general, a sample generating a signal that is higher than the cut-off value determined by this method is considered positive for a cancer.

In a related embodiment, the assay is performed in a flow-through or strip test format, wherein the binding agent is immobilized on a membrane, such as nitrocellulose. In the flow-through test, polypeptides within the sample bind to the immobilized binding agent as the sample passes through the membrane. A second, labeled binding agent then binds to the binding agent-polypeptide complex as a solution containing the second binding agent flows through the membrane. The detection of bound second binding agent may then be performed as described above. In the strip test format, one end of the membrane to which binding agent is bound is immersed in a solution containing the sample. The sample migrates along the membrane through a region containing second binding agent and to the area of immobilized binding agent. Concentration of second binding agent at the area of immobilized antibody indicates the presence of a cancer. Typically, the concentration of second binding agent at that site generates a pattern, such as a line, that can be read visually. The absence of such a pattern indicates a negative result. In general, the amount of binding agent immobilized on the membrane is selected to generate a visually discernible pattern when the biological sample contains a level of polypeptide that would be sufficient to generate a positive signal in the two-antibody sandwich assay, in the format discussed above. Preferred binding agents for use in such assays are antibodies and antigen-binding fragments thereof. Preferably, the amount of antibody immobilized on the membrane ranges from about

25 ng to about 1μg, and more preferably from about 50 ng to about 500 ng. Such tests can typically be performed with a very small amount of biological sample.

Of course, numerous other assay protocols exist that are suitable for use with the tumor proteins or binding agents of the present invention. The above descriptions are intended to be exemplary only. For example, it will be apparent to those of ordinary skill in the art that the above protocols may be readily modified to use colon tumor polypeptides to detect antibodies that bind to such polypeptides in a biological sample. The detection of such colon tumor protein specific antibodies may correlate with the presence of a cancer.

A cancer may also, or alternatively, be detected based on the presence of T cells that specifically react with a colon tumor protein in a biological sample. Within certain methods, a biological sample comprising $CD4^+$ and/or $CD8^-$ T cells isolated from a patient is incubated with a colon tumor polypeptide, a polynucleotide encoding such a polypeptide and/or an APC that expresses at least an immunogenic portion of such a polypeptide, and the presence or absence of specific activation of the T cells is detected. Suitable biological samples include, but are not limited to, isolated T cells. For example, T cells may be isolated from a patient by routine techniques (such as by Ficoll/Hypaque density gradient centrifugation of peripheral blood lymphocytes). T cells may be incubated *in vitro* for 2-9 days (typically 4 days) at 37°C with one or more representative polypeptides (*e.g.*, 5 - 25 μg/ml). It may be desirable to incubate another aliquot of a T cell sample in the absence of colon tumor polypeptide to serve as a control. For $CD4^+$ T cells, activation is preferably detected by evaluating proliferation of the T cells. For $CD8^+$ T cells, activation is preferably detected by evaluating cytolytic activity. A level of proliferation that is at least two fold greater and/or a level of cytolytic activity that is at least 20% greater than in disease-free patients indicates the presence of a cancer in the patient.

As noted above, a cancer may also, or alternatively, be detected based on the level of mRNA encoding a colon tumor protein in a biological sample. For example, at least two oligonucleotide primers may be employed in a polymerase chain reaction (PCR) based assay to amplify a portion of a colon tumor cDNA derived from a biological sample, wherein at least one of the oligonucleotide primers is specific for (*i.e.*, hybridizes to) a polynucleotide encoding the colon tumor protein. The amplified cDNA is then separated and detected using techniques well known in the art, such as gel electrophoresis. Similarly, oligonucleotide

probes that specifically hybridize to a polynucleotide encoding a colon tumor protein may be used in a hybridization assay to detect the presence of polynucleotide encoding the tumor protein in a biological sample.

To permit hybridization under assay conditions, oligonucleotide primers and probes should comprise an oligonucleotide sequence that has at least about 60%, preferably at least about 75% and more preferably at least about 90%, identity to a portion of a polynucleotide encoding a colon tumor protein that is at least 10 nucleotides, and preferably at least 20 nucleotides, in length. Preferably, oligonucleotide primers and/or probes will hybridize to a polynucleotide encoding a polypeptide disclosed herein under moderately stringent conditions, as defined above. Oligonucleotide primers and/or probes which may be usefully employed in the diagnostic methods described herein preferably are at least 10-40 nucleotides in length. In a preferred embodiment, the oligonucleotide primers comprise at least 10 contiguous nucleotides, more preferably at least 15 contiguous nucleotides, of a DNA molecule having a sequence recited in SEQ ID NO: 1-120. Techniques for both PCR based assays and hybridization assays are well known in the art (*see*, for example, Mullis et al., *Cold Spring Harbor Symp. Quant. Biol.*, *51*:263, 1987; Erlich ed., *PCR Technology*, Stockton Press, NY, 1989).

One preferred assay employs RT-PCR, in which PCR is applied in conjunction with reverse transcription. Typically, RNA is extracted from a biological sample, such as biopsy tissue, and is reverse transcribed to produce cDNA molecules. PCR amplification using at least one specific primer generates a cDNA molecule, which may be separated and visualized using, for example, gel electrophoresis. Amplification may be performed on biological samples taken from a test patient and from an individual who is not afflicted with a cancer. The amplification reaction may be performed on several dilutions of cDNA spanning two orders of magnitude. A two-fold or greater increase in expression in several dilutions of the test patient sample as compared to the same dilutions of the non-cancerous sample is typically considered positive.

In another embodiment, the disclosed compositions may be used as markers for the progression of cancer. In this embodiment, assays as described above for the diagnosis of a cancer may be performed over time, and the change in the level of reactive polypeptide(s) or polynucleotide evaluated. For example, the assays may be performed every

24-72 hours for a period of 6 months to 1 year, and thereafter performed as needed. In general, a cancer is progressing in those patients in whom the level of polypeptide or polynucleotide detected increases over time. In contrast, the cancer is not progressing when the level of reactive polypeptide or polynucleotide either remains constant or decreases with time.

Certain *in vivo* diagnostic assays may be performed directly on a tumor. One such assay involves contacting tumor cells with a binding agent. The bound binding agent may then be detected directly or indirectly via a reporter group. Such binding agents may also be used in histological applications. Alternatively, polynucleotide probes may be used within such applications.

As noted above, to improve sensitivity, multiple colon tumor protein markers may be assayed within a given sample. It will be apparent that binding agents specific for different proteins provided herein may be combined within a single assay. Further, multiple primers or probes may be used concurrently. The selection of tumor protein markers may be based on routine experiments to determine combinations that results in optimal sensitivity. In addition, or alternatively, assays for tumor proteins provided herein may be combined with assays for other known tumor antigens.

DIAGNOSTIC KITS

The present invention further provides kits for use within any of the above diagnostic methods. Such kits typically comprise two or more components necessary for performing a diagnostic assay. Components may be compounds, reagents, containers and/or equipment. For example, one container within a kit may contain a monoclonal antibody or fragment thereof that specifically binds to a colon tumor protein. Such antibodies or fragments may be provided attached to a support material, as described above. One or more additional containers may enclose elements, such as reagents or buffers, to be used in the assay. Such kits may also, or alternatively, contain a detection reagent as described above that contains a reporter group suitable for direct or indirect detection of antibody binding.

Alternatively, a kit may be designed to detect the level of mRNA encoding a colon tumor protein in a biological sample. Such kits generally comprise at least one oligonucleotide probe or primer, as described above, that hybridizes to a polynucleotide

encoding a colon tumor protein. Such an oligonucleotide may be used, for example, within a PCR or hybridization assay. Additional components that may be present within such kits include a second oligonucleotide and/or a diagnostic reagent or container to facilitate the detection of a polynucleotide encoding a colon tumor protein.

The following Examples are offered by way of illustration and not by way of limitation.

EXAMPLES

Example 1

ISOLATION AND CHARACTERIZATION OF COLON TUMOR POLYPEPTIDES BY PCR-BASED SUBTRACTION AND MICROARRAY ANALYSIS

A cDNA library was constructed in the PCR2.1 vector (Invitrogen, Carlsbad, CA) by subtracting a pool of three colon tumors with a pool of normal colon, spleen, brain, liver, kidney, lung, stomach and small intestine using PCR subtraction methodologies (Clontech, Palo Alto, CA). The subtraction was performed using a PCR-based protocol, which was modified to generate larger fragments. Within this protocol, tester and driver double stranded cDNA were separately digested with five restriction enzymes that recognize six-nucleotide restriction sites (MluI, MscI, PvuII, SalI and StuI). This digestion resulted in an average cDNA size of 600 bp, rather than the average size of 300 bp that results from digestion with RsaI according to the Clontech protocol. This modification did not affect the subtraction efficiency. Two tester populations were then created with different adapters, and the driver library remained without adapters.

The tester and driver libraries were then hybridized using excess driver cDNA. In the first hybridization step, driver was separately hybridized with each of the two tester cDNA populations. This resulted in populations of (a) unhybridized tester cDNAs, (b) tester cDNAs hybridized to other tester cDNAs, (c) tester cDNAs hybridized to driver cDNAs and (d) unhybridized driver cDNAs. The two separate hybridization reactions were then combined, and rehybridized in the presence of additional denatured driver cDNA. Following this second hybridization, in addition to populations (a) through (d), a fifth population (e) was generated in which tester cDNA with one adapter hybridized to tester cDNA with the second adapter. Accordingly, the second hybridization step resulted in enrichment of differentially expressed sequences which could be used as templates for PCR amplification with adaptor-specific primers.

The ends were then filled in, and PCR amplification was performed using adaptor-specific primers. Only population (e), which contained tester cDNA that did not hybridize to driver cDNA, was amplified exponentially. A second PCR amplification step was then performed, to reduce background and further enrich differentially expressed sequences.

This PCR-based subtraction technique normalizes differentially expressed cDNAs so that rare transcripts that are overexpressed in prostate tumor tissue may be recoverable. Such transcripts would be difficult to recover by traditional subtraction methods.

To characterize the complexity and redundancy of the subtracted library, ninety six clones were randomly picked and sixty five were sequenced, as previously described. These sequences were further characterized by comparison with the most recent Genbank database (April, 1998) to determine their degree of novelty. No significant homologies were found to twenty one of these clones, hereinafter referred to as 11092, 11093, 11096, 11098, 11103, 11174, 11108, 11112, 11115, 11117, 11118, 11134, 11151, 11154, 11158, 11168, 11172, 11175, 11184, 11185 and 11187. The determined cDNA sequences for these clones are provided in SEQ ID NO: 48, 49, 52, 54, 59, 60, 65-69, 79, 89, 90, 93, 99-101 and 109-111, respectively.

Two thousand clones from the above mentioned cDNA subtraction library were randomly picked and submitted to a round of PCR amplification. Briefly, 0.5 µl of glycerol stock solution was added to 99.5 µl of pcr MIX (80 µl H_20, 10 µl 10X PCR Buffer, 6 µl 25 mM $MgCl_2$, 1 µl 10 mM dNTPs, 1 µl 100 mM M13 forward primer (CACGACGTTGTAAAACGACGG), 1 µl 100 mM M13 reverse primer (CACAGGAAACAGCTATGACC), and 0.5 µl 5 u/ml Taq polymerase (primers provided by (Operon Technologies, Alameda, CA). The PCR amplification was run for thirty cycles under the following conditions: 95°C for 5 min., 92°C for 30 sec., 57°C for 40 sec., 75°C for 2 min. and 75°C for 5 minutes.

mRNA expression levels for representative clones were determined using microarray technology (Synteni, Palo Alto, CA) in colon tumor tissues (n=25), normal colon tissues (n=6), kidney, lung, liver, brain, heart, esophagus, small intestine, stomach, pancreas, adrenal gland, salivary gland, resting PBMC, activated PBMC, bone marrow, dendritic cells, spinal

cord, blood vessels, skeletal muscle, skin, breast and fetal tissues. The number of tissue samples tested in each case was one (n=1), except where specifically noted above; additionally, all the above-mentioned tissues were derived from human. The PCR amplification products were dotted onto slides in an array format, with each product occupying a unique location in the array. mRNA was extracted from the tissue sample to be tested, and fluorescent-labeled cDNA probes were generated by reverse transcription according to the protocol provided by Synteni. The microarrays were probed with the labeled cDNA probes, the slides scanned, and fluorescence intensity was measured. This intensity correlates with the hybridization intensity.

One hundred and forty nine clones showed two or more fold over-expression in the colon tumor probe group as compared to the normal tissue probe group. These cDNA clones were further characterized by DNA sequencing with a Perkin Elmer/Applied Biosystems Division Automated Sequencer Model 373A and/or Model 377 (Foster City, CA). These sequences were compared to known sequences in the most recent GenBank database. No significant homologies to human gene sequences were found in forty nine of these clones, represented by the following sixteen cDNA consensus sequences: SEQ ID NO: 2, 8, 15, 16, 22, 24, 30, 32-34, 36, 38, 40, 41, 46 and 47, hereinafter referred to as Contig 2, 8, 13, 14, 20, 23, 29, 31, 35, 32, 36, 38, 41, 42, 50 and 51, respectively). Contig 29 (SEQ ID NO: 30) was found to be a Rat GSK-3-β-interacting protein Axil homolog. Also, Contigs 31 and 35 (SEQ ID NO: 32 and 33, respectively) were found to be a Mus musculus GOB-4 homolog. The determined cDNA sequences of SEQ ID NO: 1, 3-7, 9-14, 17-21, 23, 25-29, 31, 35, 37, 39, 42-45, 50, 51, 53, 55-58, 61-64, 70-78, 80-88, 91, 92, 94-98, 102-108 and 112 were found to show some homology to previously identified genes sequences.

Microarray analysis demonstrated Contig 2 (SEQ ID NO: 2) showed over-expression in 34% of colon tumors tested, as well as increased expression in normal pancreatic tissue, with no over-expression in normal colon tissues. Upon further analysis, Contigs 2, 8 and 23 were found to share homology to the known gene GW112. Contigs 4, 5, 9 and 52 showed homology to carcinoembryonic antigen (SEQ ID NO: 3, 4, 5 and 6, respectively). A representative sampling of these fragments showed over-expression in 85% of colon tumors, with over-expression in normal bone marrow and 3/6 normal colon tissues. Contig 6 (SEQ ID NO: 7), showing homology to the known gene sequence for villin, and was overexpressed

in about half of all colon tumors tested, with a limited degree of low level over-expression in normal colon. Contig 12 (SEQ ID NO: 14), showing homology to Chromosome 17, clone hRPC.1171_I_10, also referred to as C798P, was over-expressed in approximately 70% of colon tumors tested, with low over-expression in 1/6 normal colon samples. Contig 14, also referred to as 14261 (SEQ ID NO: 16), showing no significant homology to any known gene, showed over-expression in 44% of colon tumors tested, with low level expression in half of normal colon tissues, as well as small intestine and pancreatic tissue. Contig 18 (SEQ ID NO: 21), showing homology to the known gene for L1-cadherin showed over-expression in approximately half of colon tumors and low level overespression in 3/6 normal colon tissues tested. Contig 22 (SEQ ID NO: 23), showing homology to Bumetanide-sensitive Na-K-Cl cotransporter was over-expressed in 70% of colon tumors and no over-expression in all normal tissues tested. Contig 25 (SEQ ID NO: 25), showing homology to macrophage inflammatory protein-3α, was over-expressed in over 40% of colon tumors and in activated PBMC. Contigs 26 and 48 (SEQ ID NOS: 25 and 26), showing homology to the sequence for laminin, was over-expressed in 48% of colon tumors and with low over-expression in stomach tissue. Contig 28 (SEQ ID NO: 29), showing homology to the known gene sequence for Chromosome 16 BAC clone CIT987SK-A-363E6, was over-expressed in 33% of colon tumors tested with normal stomach and 2/6 normal colon tissues showing low level over-expression. Contigs 29, 31 and 35 (SEQ ID NOS: 30, 32 and 33, respetively), also referred to as C751P, an unknown sequence showing limited and partial homology to Rat GSK-3β-interacting protein Axil homolog.and Mus musculus GOB-4 homolog, was over-expressed in 74% of colon tumors and no over-expression in all normal tissues tested. Contig 34 (SEQ ID NO: 35), showing homology to the known sequence for desmoglein 2, was over-expressed in 56% of colon tumors and showed low level over-expression in 1/6 normal colon tissues. Contig 36 (SEQ ID NO: 36), an unknown sequence also referred to as C793P, showed over-expression in 30% of colon tumor tissues tested. Contig 37 and 14287.2 (SEQ ID NOS: 37 and 116), an unknown sequence, but with limited (89%) homology to the known sequence for putative transmembrane protein was over-expressed in 70% of colon tumors, as well as in normal lung tissue and 3/6 normal colon tissues tested. Contig 38, also referred to as C796P and 14219 (SEQ ID NO: 38), showing no significant homology to any known gene, was over-expressed in 38% in colon tumors and no elevated over-expression in any normal

tissues. Contig 41 (SEQ ID NO: 40), also referred to as C799P and 14308, an unknown sequence showing no significant homology to any known gene, was over-expressed in 22% of colon tumors. Contig 42, (SEQ ID NO: 41), also referred to as C794P and 14309, an unknown sequence with no significant homology to any known gene, was over-expressed in 63% of colon tumors tested, as well as in 3/6 normal colon tissues. Contig 43 (SEQ ID NO: 42), showing homology to the known sequence for Chromosome 1 specific transcript KIAA0487 was over-expressed in 85% of colon tumors tested and in normal lung and 4/6 normal colon tissues. Contig 49 (SEQ ID NO: 45), showing homology to the known sequence for pump-1, was over-expressed in 44% of colon tumors and no over-expression in all normal tissues tested. Contig 50 (SEQ ID NO: 46), also referred to as C792P and 18323, showing no significant homology to any known gene, was over-expressed in 33% of colon tumors with no detectable over-expression in any normal tissues tested. Contig 51 (SEQ ID NO: 47), also referred to as C795P and 14317 was over-expressed in 11% of colon tumors. To the best of the inventors' knowledge, none of these sequences have been previously shown to be present in colon.

Additional microarray analysis yielded seven clones showing two or more fold over-expression in the colon tumor probe group as compared to the normal tissue probe group. Three of these clones demonstrated particularly good colon tumor specificity, and are represented by SEQ ID NO: 115, 116 and 120. Specifically, SEQ ID NO: 115, referred to as C791P and 14235, which shows homology to the known gene sequence for H. sapiens chromosome 21 derived BAC containing ets-2 gene, was over-expressed in 89% of colon tumors tested and in 5/6 normal colon tissues, as well as over-expressed at low levels in normal lung and activated PBMC. Microarray analysis for SEQ ID NO: 116 is discussed above. SEQ ID NO: 120, referred to as 14295, showing homology to the known gene sequence for secreted cement gland protein XAG-2 homolog, was over-expressed in 70% of colon tumors and in 5/6 normal colon tissues, as well as low level over-expression in normal small intestine, stomach and lung. All clones showing over-expression in colon tumor were sequenced and these sequences compared to the most recent Genbank database (February 12, 1999). Of the seven clones, three contained sequences that did not share significant homology to any known gene sequences, represented by SEQ ID NO: 116, 117 and 119. To the best of the inventors' knowledge, none of these sequences have been previously shown to

be present in colon. The determined cDNA sequences of the remaining clones (SEQ ID NO: 113-115 and 120) were found to show significant homology to previously identified genes.

Further analysis revealed a clone which was recovered several times by PCR subtraction and by expression screening using a mouse anti-scid antiserum. The determined full length cDNA sequence, together with the known amino acid sequence for P733P are provided in Figures 1A-D. Clone P733P is homologous with the known gene Beta IG-H3, as disclosed in U.S. Patent No. 5,444,164. Microarray analysis demonstrated this clone to be over-expressed in 75 to 80% of colon tumors tested (n=27), with no over-expression in normal colon samples (n=6), but with some low level over-expression in other normal tissues tested.

Example 2
SYNTHESIS OF POLYPEPTIDES

Polypeptides may be synthesized on a Perkin Elmer/Applied Biosystems Division 430A peptide synthesizer using FMOC chemistry with HPTU (O-Benzotriazole-N,N,N',N'-tetramethyluronium hexafluorophosphate) activation. A Gly-Cys-Gly sequence may be attached to the amino terminus of the peptide to provide a method of conjugation, binding to an immobilized surface, or labeling of the peptide. Cleavage of the peptides from the solid support may be carried out using the following cleavage mixture: trifluoroacetic acid:ethanedithiol:thioanisole:water:phenol (40:1:2:2:3). After cleaving for 2 hours, the peptides may be precipitated in cold methyl-t-butyl-ether. The peptide pellets may then be dissolved in water containing 0.1% trifluoroacetic acid (TFA) and lyophilized prior to purification by C18 reverse phase HPLC. A gradient of 0%-60% acetonitrile (containing 0.1% TFA) in water (containing 0.1% TFA) may be used to elute the peptides. Following lyophilization of the pure fractions, the peptides may be characterized using electrospray or other types of mass spectrometry and by amino acid analysis.

From the foregoing it will be appreciated that, although specific embodiments of the invention have been described herein for purposes of illustration, various modifications

may be made without deviating from the spirit and scope of the invention. Accordingly, the invention is not limited except as by the appended claims.

CLAIMS

1. An isolated polypeptide comprising at least an immunogenic portion of a colon tumor protein, or a variant thereof, wherein the tumor protein comprises an amino acid sequence that is encoded by a polynucleotide sequence selected from the group consisting of:

(a) sequences recited in SEQ ID NO: 2, 8, 15, 16, 22, 24, 30, 32-34, 36, 38, 40, 41, 46-49, 52, 54, 59, 60, 65-69, 79, 89, 90, 93, 99-101, 109-111 or 116-119;

(b) sequences that hybridize to a sequence of SEQ ID NO: 2, 8, 15, 16, 22, 24, 30, 32-34, 36, 38, 40, 41, 46-49, 52, 54, 59, 60, 65-69, 79, 89, 90, 93, 99-101, 109-111 or 116-119 under moderately stringent conditions; and

(c) a complement of a sequence of (a) or (b).

2. An isolated polypeptide according to claim 1, wherein the polypeptide comprises an amino acid sequence that is encoded by a polynucleotide sequence recited in any one of SEQ ID NO: 2, 8, 15, 16, 22, 24, 30, 32-34, 36, 38, 40, 41, 46-49, 52, 54, 59, 60, 65-69, 79, 89, 90, 93, 99-101, 109-111, 116-119 or a complement of any of the foregoing polynucleotide sequences.

3. An isolated polypeptide encoding at least 15 amino acid residues of a colon tumor protein, or a variant thereof that differs in one or more substitutions, deletions, additions and/or insertions such that the ability of the variant to react with antigen-specific antisera is not substantially diminished, wherein the tumor protein comprises an amino acid sequence that is encoded by a polynucleotide comprising a sequence recited in any one of SEQ ID NO: 2, 8, 15, 16, 22, 24, 30, 32-34, 36, 38, 40, 41, 46-49, 52, 54, 59, 60, 65-69, 79, 89, 90, 93, 99-101, 109-111 or 116-119 or a complement of any of the foregoing sequences.

4. An isolated polynucleotide encoding a colon tumor protein, or a variant thereof, wherein the tumor protein comprises an amino acid sequence that is encoded by a polynucleotide comprising a sequence recited in any one of SEQ ID NO: 2, 8, 15, 16, 22, 24,

30, 32-34, 36, 38, 40, 41, 46-49, 52, 54, 59, 60, 65-69, 79, 89, 90, 93, 99-101, 109-111 and 116-119 or a complement of any of the foregoing sequences.

5. An isolated polynucleotide comprising a sequence recited in any one of SEQ ID NO: 2, 8, 15, 16, 22, 24, 30, 32-34, 36, 38, 40, 41, 46-49, 52, 54, 59, 60, 65-69, 79, 89, 90, 93, 99-101, 109-111 and 116-119.

6. An isolated polynucleotide comprising a sequence that hybridizes to a sequence recited in any one of SEQ ID NO: 2, 8, 15, 16, 22, 24, 30, 32-34, 36, 38, 40, 41, 46-49, 52, 54, 59, 60, 65-69, 79, 89, 90, 93, 99-101, 109-111 and 116-119 under moderately stringent conditions.

7. An isolated polynucleotide complementary to a polynucleotide according to any one of claims 3-6.

8. An expression vector comprising a polynucleotide according to any one of claims claim 3-6.

9. A host cell transformed or transfected with an expression vector according to claim 8.

10. An expression vector comprising a polynucleotide according claim 7.

11. A host cell transformed or transfected with an expression vector according to claim 10.

12. A pharmaceutical composition comprising a polypeptide according to claim 1, in combination with a physiologically acceptable carrier.

13. A vaccine comprising a polypeptide according to claim 1, in combination with a non-specific immune response enhancer.

14. A vaccine according to claim 13, wherein the non-specific immune response enhancer is an adjuvant.

15. A vaccine according to claim 13, wherein the non-specific immune response enhancer induces a predominantly Type I response.

16. A pharmaceutical composition comprising a polynucleotide according to claim 3, in combination with a physiologically acceptable carrier.

17. A vaccine comprising a polynucleotide according to claim 3, in combination with a non-specific immune response enhancer.

18. A vaccine according to claim 17, wherein the non-specific immune response enhancer is an adjuvant.

19. A vaccine according to claim 17, wherein the non-specific immune response enhancer induces a predominantly Type I response.

20. An isolated antibody, or antigen-binding fragment thereof, that specifically binds to a colon tumor protein that comprises an amino acid sequence that is encoded by a polynucleotide sequence recited in any one of SEQ ID NO: 2, 8, 15, 16, 22, 24, 30, 32-34, 36, 38, 40, 41, 46-49, 52, 54, 59, 60, 65-69, 79, 89, 90, 93, 99-101, 109-111 and 116-119 or a complement of any of the foregoing polynucleotide sequences.

21. A pharmaceutical composition comprising an antibody or fragment thereof according to claim 17, in combination with a physiologically acceptable carrier.

22. A pharmaceutical composition comprising an antigen-presenting cell that expresses a polypeptide according to claim 1, in combination with a pharmaceutically acceptable carrier or excipient.

23. A pharmaceutical composition according to claim 22, wherein the antigen presenting cell is a dendritic cell or a macrophage.

24. A vaccine comprising an antigen-presenting cell that expresses a polypeptide according to claim 1, in combination with a non-specific immune response enhancer.

25. A vaccine according to claim 24, wherein the non-specific immune response enhancer is an adjuvant.

26. A vaccine according to claim 24, wherein the non-specific immune response enhancer induces a predominantly Type I response.

27. A vaccine according to claim 24, wherein the antigen-presenting cell is a dendritic cell.

28. A method for inhibiting the development of a cancer in a patient, comprising administering to a patient an effective amount of a polypeptide according to claim 1, and thereby inhibiting the development of a cancer in the patient.

29. A method for inhibiting the development of a cancer in a patient, comprising administering to a patient an effective amount of a polynucleotide according to claim 3, and thereby inhibiting the development of a cancer in the patient.

30. A method for inhibiting the development of a cancer in a patient, comprising administering to a patient an effective amount of an antibody or antigen-binding fragment thereof according to claim 20, and thereby inhibiting the development of a cancer in the patient.

31. A method for inhibiting the development of a cancer in a patient, comprising administering to a patient an effective amount of an antigen-presenting cell that expresses a polypeptide according to claim 1, and thereby inhibiting the development of a cancer in the patient.

32. A method according to claim 31, wherein the antigen-presenting cell is a dendritic cell.

33. A method according to any one of claims 28-31, wherein the cancer is colon cancer.

34. A fusion protein comprising at least one polypeptide according to claim 1.

35. A fusion protein according to claim 34, wherein the fusion protein comprises an expression enhancer that increases expression of the fusion protein in a host cell transfected with a polynucleotide encoding the fusion protein.

36. A fusion protein according to claim 34, wherein the fusion protein comprises a T helper epitope that is not present within the polypeptide of claim 1.

37. A fusion protein according to claim 34, wherein the fusion protein comprises an affinity tag.

38. An isolated polynucleotide encoding a fusion protein according to claim 34.

39. A pharmaceutical composition comprising a fusion protein according to claim 31, in combination with a physiologically acceptable carrier.

40. A vaccine comprising a fusion protein according to claim 34, in combination with a non-specific immune response enhancer.

41. A vaccine according to claim 40, wherein the non-specific immune response enhancer is an adjuvant.

42. A vaccine according to claim 40, wherein the non-specific immune response enhancer induces a predominantly Type I response.

43. A pharmaceutical composition comprising a polynucleotide according to claim 3, in combination with a physiologically acceptable carrier.

44. A vaccine comprising a polynucleotide according to claim 3, in combination with a non-specific immune response enhancer.

45. A vaccine according to claim 44, wherein the non-specific immune response enhancer is an adjuvant.

46. A vaccine according to claim 44, wherein the non-specific immune response enhancer induces a predominantly Type I response.

47. A method for inhibiting the development of a cancer in a patient, comprising administering to a patient an effective amount of a pharmaceutical composition according to any one of claims 39 and 43.

48. A method for inhibiting the development of a cancer in a patient, comprising administering to a patient an effective amount of a vaccine according to any one of claims 40 and 44.

49. A method for removing tumor cells from a biological sample, comprising contacting a biological sample with T cells that specifically react with a colon

tumor protein, wherein the tumor protein comprises an amino acid sequence that is encoded by a polynucleotide sequence selected from the group consisting of:

(i) polynucleotides recited in any one of SEQ ID NO: 2, 8, 15, 16, 22, 24, 30, 32-34, 36, 38, 40, 41, 46-49, 52, 54, 59, 60, 65-69, 79, 89, 90, 93, 99-101, 109-111 and 116-119; and

(ii) complements of the foregoing polynucleotides;

wherein the step of contacting is performed under conditions and for a time sufficient to permit the removal of cells expressing the antigen from the sample.

50. A method according to claim 49, wherein the biological sample is blood or a fraction thereof.

51. A method for inhibiting the development of a cancer in a patient, comprising administering to a patient a biological sample treated according to the method of claim 49.

52. A method for stimulating and/or expanding T cells specific for a colon tumor protein, comprising contacting T cells with one or more of:

(i) a polypeptide according to claim 1;

(ii) a polynucleotide encoding such a polypeptide; and/or

(iii) an antigen presenting cell that expresses such a polypeptide;

under conditions and for a time sufficient to permit the stimulation and/or expansion of T cells.

53. An isolated T cell population, comprising T cells prepared according to the method of claim 52.

54. A method for inhibiting the development of a cancer in a patient, comprising administering to a patient an effective amount of a T cell population according to claim 53.

55. A method for inhibiting the development of a cancer in a patient, comprising the steps of:

(a) incubating $CD4^+$ and/or CD8+ T cells isolated from a patient with at least one component selected from the group consisting of:

(i) a polypeptide according to claim 1;

(ii) a polynucleotide encoding such a polypeptide; or

(iii) an antigen-presenting cell that expresses such a polypeptide;

such that T cells proliferate; and

(b) administering to the patient an effective amount of the proliferated T cells, and thereby inhibiting the development of a cancer in the patient.

56. A method for inhibiting the development of a cancer in a patient, comprising the steps of:

(a) incubating $CD4^+$ and/or CD8+ T cells isolated from a patient with at least one component selected from the group consisting of:

(i) a polypeptide according to claim 1;

(ii) a polynucleotide encoding such a polypeptide; or

(iii) an antigen-presenting cell that expresses such a polypeptide;

such that T cells proliferate;

(b) cloning at least one proliferated cell; and

(c) administering to the patient an effective amount of the cloned T cells, and thereby inhibiting the development of a cancer in the patient.

57. A method for determining the presence or absence of a cancer in a patient, comprising the steps of:

(a) contacting a biological sample obtained from a patient with a binding agent that binds to a colon tumor protein, wherein the tumor protein comprises an amino acid sequence that is encoded by a polynucleotide sequence selected from the group consisting of:

(i) polynucleotides recited in any one of SEQ ID NO: 1-120; and

(ii) complements of the foregoing polynucleotides;

(b) detecting in the sample an amount of polypeptide that binds to the binding agent; and

(c) comparing the amount of polypeptide to a predetermined cut-off value, and therefrom determining the presence or absence of a cancer in the patient.

58. A method according to claim 57, wherein the binding agent is an antibody.

59. A method according to claim 57, wherein the antibody is a monoclonal antibody.

60. A method according to claim 57, wherein the cancer is colon cancer.

61. A method for monitoring the progression of a cancer in a patient, comprising the steps of:

(a) contacting a biological sample obtained from a patient at a first point in time with a binding agent that binds to a colon tumor protein, wherein the tumor protein comprises an amino acid sequence that is encoded by a polynucleotide sequence recited in any one of SEQ ID NO: 1-120 or a complement of any of the foregoing polynucleotides;

(b) detecting in the sample an amount of polypeptide that binds to the binding agent;

(c) repeating steps (a) and (b) using a biological sample obtained from the patient at a subsequent point in time; and

(d) comparing the amount of polypeptide detected in step (c) to the amount detected in step (b) and therefrom monitoring the progression of the cancer in the patient.

62. A method according to claim 61, wherein the binding agent is an antibody.

63. A method according to claim 62, wherein the antibody is a monoclonal antibody.

64. A method according to claim 61, wherein the cancer is a colon cancer.

65. A method for determining the presence or absence of a cancer in a patient, comprising the steps of:

(a) contacting a biological sample obtained from a patient with an oligonucleotide that hybridizes to a polynucleotide that encodes a colon tumor protein, wherein the tumor protein comprises an amino acid sequence that is encoded by a polynucleotide sequence recited in any one of SEQ ID NO: 1-120 or a complement of any of the foregoing polynucleotides;

(b) detecting in the sample an amount of a polynucleotide that hybridizes to the oligonucleotide; and

(c) comparing the amount of polynucleotide that hybridizes to the oligonucleotide to a predetermined cut-off value, and therefrom determining the presence or absence of a cancer in the patient.

66. A method according to claim 65, wherein the amount of polynucleotide that hybridizes to the oligonucleotide is determined using a polymerase chain reaction.

67. A method according to claim 65, wherein the amount of polynucleotide that hybridizes to the oligonucleotide is determined using a hybridization assay.

68. A method for monitoring the progression of a cancer in a patient, comprising the steps of:

(a) contacting a biological sample obtained from a patient with an oligonucleotide that hybridizes to a polynucleotide that encodes a colon tumor protein, wherein the tumor protein comprises an amino acid sequence that is encoded by a polynucleotide sequence recited in any one of SEQ ID NO: 1-120 or a complement of any of the foregoing polynucleotides;

(b) detecting in the sample an amount of a polynucleotide that hybridizes to the oligonucleotide;

(c) repeating steps (a) and (b) using a biological sample obtained from the patient at a subsequent point in time; and

(d) comparing the amount of polynucleotide detected in step (c) to the amount detected in step (b) and therefrom monitoring the progression of the cancer in the patient.

69. A method according to claim 68, wherein the amount of polynucleotide that hybridizes to the oligonucleotide is determined using a polymerase chain reaction.

70. A method according to claim 68, wherein the amount of polynucleotide that hybridizes to the oligonucleotide is determined using a hybridization assay.

71. A diagnostic kit, comprising:

(a) one or more antibodies according to claim 20; and

(b) a detection reagent comprising a reporter group.

72. A kit according to claim 71, wherein the antibodies are immobilized on a solid support.

73. A kit according to claim 72, wherein the solid support comprises nitrocellulose, latex or a plastic material.

74. A kit according to claim 71, wherein the detection reagent comprises an anti-immunoglobulin, protein G, protein A or lectin.

75. A kit according to claim 71, wherein the reporter group is selected from the group consisting of radioisotopes, fluorescent groups, luminescent groups, enzymes, biotin and dye particles.

76. An oligonucleotide comprising 10 to 40 nucleotides that hybridize under moderately stringent conditions to a polynucleotide that encodes a colon tumor protein,

wherein the tumor protein comprises an amino acid sequence that is encoded by a polynucleotide sequence recited in any one of SEQ ID NO: 2, 8, 15, 16, 22, 24, 30, 32-34, 36, 38, 40, 41, 46-49, 52, 54, 59, 60, 65-69, 79, 89, 90, 93, 99-101, 109-111 and 116-119 or a complement of any of the foregoing polynucleotides.

77. A oligonucleotide according to claim 76, wherein the oligonucleotide comprises 10-40 nucleotides recited in any one of SEQ ID NO: 2, 8, 15, 16, 22, 24, 30, 32-34, 36, 38, 40, 41, 46-49, 52, 54, 59, 60, 65-69, 79, 89, 90, 93, 99-101, 109-111 and 116-119.

78. A diagnostic kit, comprising:

(a) an oligonucleotide according to claim 76; and

(b) a diagnostic reagent for use in a polymerase chain reaction or hybridization assay.

COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS OF COLON CANCER AND METHODS FOR THEIR USE

ABSTRACT OF THE DISCLOSURE

Compositions and methods for the therapy and diagnosis of cancer, such as colon cancer, are disclosed. Compositions may comprise one or more colon tumor proteins, immunogenic portions thereof, or polynucleotides that encode such portions. Alternatively, a therapeutic composition may comprise an antigen presenting cell that expresses a colon tumor protein, or a T cell that is specific for cells expressing such a protein. Such compositions may be used, for example, for the prevention and treatment of diseases such as colon cancer. Diagnostic methods based on detecting a colon tumor protein, or mRNA encoding such a protein, in a sample are also provided.

<210> 1
<211> 2691
<212> DNA
<213> Homo sapiens

<400> 1
gcttgcccgt cggtcgctag ctcgctcggt gcgcgtcgtc ccgctccatg grgctcttcg 60
tgcggctgct ggctctcgcc ctggctctgg ccctgggccc cgccgcgacc ctggcgggtc 120
ccgccaagtc gccctaccag ctggtgctgc agcacagcag gctccggggc cgccagcacg 180
gccccaacgt gtgtgctgtg cagaaggtta ttggcactaa taggaagtac ttcaccaact 240
gcaagcagtg gtaccaaagg aaaatctgtg gcaaatcaac agtcatcagc tacgagtgct 300
gtcctggata tgaaaaggtc cctggggaga agggctgtcc agcagcccta ccactctcaa 360
accttacga gaccctggga gtcgttggat ccaccaccac tcagctgtac acggaccgca 420
cggagaagct gaggcctgag atggagggcc ccggcagctt caccatcttc gccctagca 480
acgaggcctg ggcctccttg ccagctgaag tgctggactc cctggtcagc aatgtcaaca 540
ttgagctgct caatgccctc cgctaccata tggtgggcag gcgagtcctg actgatgagc 600
tgaaacacgg catgaccctc acctctatgt accagaattc caacatccag atccaccact 660
atcctaatgg gattgtaact gtgaactgtg ccgggctcct gaaagccgac caccatgcaa 720
ccaacgggct ggtgcaccta atcgataagg tcatctccac catcaccaac aacatccagc 780
agatcattga gatcgaggac accttcgaga tcctcgggtc tgctgtgggt gcatcagggc 840
tcaacacgat gcttgaaggt aacggccagt acacgctttt ggcccgacc aatgaggcct 900
tcgagaagat ccctagtgag acttcgaacc gtatcctggg cgacccagaa gcccctgagag 960
acctgctgaa caaccacatc ttgaagtcag ctatgtgtgc tgaagccatc gttgcgggtc 1020
tgtctgtaga gacccctggag ggcacgacac tggaggtggg ctgcagcggg gacatgctca 1080
ctatcaacgg gaaggcgatc atctccaata aagacatcct agccaccaac ggggtgatcc 1140
actacattga tgagctactc atcccagact cagccaagac actatttgaa ttggctgcag 1200
agtctgatgt gtccacagcc attgaccttt tcagacaagc cggcctcggc aatcatctct 1260
ctggaagtga gcggttgacc ctcctggctc ccctgaattc tgtattcaaa gatggaaccc 1320
ctccaattga tgcccataca aggaatttgc ttcggaacca cataattaaa gaccagctgg 1380
cctctaagta tctgtaccat ggacagacgc tggaaactct gggcggcaaa aaacttgagag 1440
tttttgttta tcgtaatagc ctctgcattg agaacagctg catcgcggcc cacgacaaga 1500
ggggaggta cgggaccctg ttcacgatgg accgggtgct gaccccccca atgggactg 1560
tcatggatgt cctgaagggga gacaatcgct ttagcatgct ggtagctgcc atccagtctg 1620
caggactgac ggagaccctc aaccgggaag gagtctacac agtctttgct cccacaaatg 1680
aagccttccg agccctgcca ccaagagaac ggagcagact cttgggagat gccaaggaac 1740
ttgccaacat cctgaaatac cacattggtg atgaaatcct ggttagcgga ggcatcgggg 1800
ccctggtgcg gctaaagtct ctccaaggtg acaagcttga agtcagcttg aaaaacaatg 1860
tggtgagtgt caacaaggag ccrgttgccg agcctgacat catggccaca aatggcgtgg 1920
tccatgtcat caccaatgtt ccgcagccct cagccaacag accttaggaa agaggggatg 1980
aacttgcaga ctctgcgctt gagatcttca aacaagcatc agcgttttcc agggcttccc 2040
agaggtctgt gcgactagcc cctgtcctatc aaaagttatt agagaggatg aagcattagc 2100
ttgaagcact acaggaggaa tgcaccacgg cagctctccg ccaatttctc tcagatttcc 2160
acagagactg tttgaatgtt ttcaaaacca agtatcacac ttcaatgtac acgggccgca 2220
ccataatgag atgtgagcct tgtgcatgtg ggggaggagg gagagagatg tacttttaa 2280
atcatgrtcc ccctaaacat ggctgttaac ccactgcatg cagaaacttg gatgtcactg 2340
cctgacattc acttccagag aggacctatc ccaaatgtgg aattgactgc ctatgccaag 2400

FIG. 1A

```
tccctggaaa aggagcttca gtattgtggg gcttcataaaa catgaatcaa gcaatccagc 2460
ctcatgggaa gtcctggcac agttttttgta aagccctttgc acagctggag aaatggcatc 2520
attataagct atgagttgaa atgttctgtc aaatgtgtct cacatctaca cgtggcttgg 2580
aggcttttat ggggccctgt ccaggtagaa aagaaatggt atgtagagct tagatttccc 2640
tactgtgaca gagccatggt gtgtttgtaa taataaaacc aaagaaacat a          2691
```

<210> 2
<211> 683
<212> PRT
<213> Homo sapiens

<400> 2

```
Met Ala Leu Phe Val Arg Leu Leu Ala Leu Ala Leu Ala Leu Ala Leu
                5                  10                  15

Gly Pro Ala Ala Thr Leu Ala Gly Pro Ala Lys Ser Pro Tyr Gln Leu
            20                  25                  30

Val Leu Gln His Ser Arg Leu Arg Gly Arg Gln His Gly Pro Asn Val
        35                  40                  45

Cys Ala Val Gln Lys Val Ile Gly Thr Asn Arg Lys Tyr Phe Thr Asn
    50                  55                  60

Cys Lys Gln Trp Tyr Gln Arg Lys Ile Cys Gly Lys Ser Thr Val Ile
65                  70                  75                  80

Ser Tyr Glu Cys Cys Pro Gly Tyr Glu Lys Val Pro Gly Glu Lys Gly
                85                  90                  95

Cys Pro Ala Ala Leu Pro Leu Ser Asn Leu Tyr Glu Thr Leu Gly Val
            100                 105                 110

Val Gly Ser Thr Thr Thr Gln Leu Tyr Thr Asp Arg Thr Glu Lys Leu
        115                 120                 125

Arg Pro Glu Met Glu Gly Pro Gly Ser Phe Thr Ile Phe Ala Pro Ser
    130                 135                 140

Asn Glu Ala Trp Ala Ser Leu Pro Ala Glu Val Leu Asp Ser Leu Val
145                 150                 155                 160

Ser Asn Val Asn Ile Glu Leu Leu Asn Ala Leu Arg Tyr His Met Val
                165                 170                 175

Gly Arg Arg Val Leu Thr Asp Glu Leu Lys His Gly Met Thr Leu Thr
            180                 185                 190

Ser Met Tyr Gln Asn Ser Asn Ile Gln Ile His His Tyr Pro Asn Gly
        195                 200                 205

Ile Val Thr Val Asn Cys Ala Arg Leu Leu Lys Ala Asp His His Ala
    210                 215                 220

Thr Asn Gly Val Val His Leu Ile Asp Lys Val Ile Ser Thr Ile Thr
225                 230                 235                 240

Asn Asn Ile Gln Gln Ile Ile Glu Ile Glu Asp Thr Phe Glu Thr Leu
                245                 250                 255

Arg Ala Ala Val Ala Ala Ser Gly Leu Asn Thr Met Leu Glu Gly Asn
            260                 265                 270

Gly Gln Tyr Thr Leu Leu Ala Pro Thr Asn Glu Ala Phe Glu Lys Ile
```

FIG. 1B

```
        275                 280                 285

Pro Ser Glu Thr Leu Asn Arg Ile Leu Gly Asp Pro Glu Ala Leu Arg
    290                 295                 300

Asp Leu Leu Asn Asn His Ile Leu Lys Ser Ala Met Cys Ala Glu Ala
305                 310                 315                 320

Ile Val Ala Gly Leu Ser Val Glu Thr Leu Glu Gly Thr Thr Leu Glu
                325                 330                 335

Val Gly Cys Ser Gly Asp Met Leu Thr Ile Asn Gly Lys Ala Ile Ile
            340                 345                 350

Ser Asn Lys Asp Ile Leu Ala Thr Asn Gly Val Ile His Tyr Ile Asp
        355                 360                 365

Glu Leu Leu Ile Pro Asp Ser Ala Lys Thr Leu Phe Glu Leu Ala Ala
    370                 375                 380

Glu Ser Asp Val Ser Thr Ala Ile Asp Leu Phe Arg Gln Ala Gly Leu
385                 390                 395                 400

Gly Asn His Leu Ser Gly Ser Glu Arg Leu Thr Leu Leu Ala Pro Leu
                405                 410                 415

Asn Ser Val Phe Lys Asp Gly Thr Pro Pro Ile Asp Ala His Thr Arg
            420                 425                 430

Asn Leu Leu Arg Asn His Ile Ile Lys Asp Gln Leu Ala Ser Lys Tyr
        435                 440                 445

Leu Tyr His Gly Gln Thr Leu Glu Thr Leu Gly Gly Lys Lys Leu Arg
    450                 455                 460

Val Phe Val Tyr Arg Asn Ser Leu Cys Ile Glu Asn Ser Cys Ile Ala
465                 470                 475                 480

Ala His Asp Lys Arg Gly Arg Tyr Gly Thr Leu Phe Thr Met Asp Arg
                485                 490                 495

Val Leu Thr Pro Pro Met Gly Thr Val Met Asp Val Leu Lys Gly Asp
            500                 505                 510

Asn Arg Phe Ser Met Leu Val Ala Ala Ile Gln Ser Ala Gly Leu Thr
        515                 520                 525

Glu Thr Leu Asn Arg Glu Gly Val Tyr Thr Val Phe Ala Pro Thr Asn
    530                 535                 540

Glu Ala Phe Arg Ala Leu Pro Pro Arg Glu Arg Ser Arg Leu Leu Gly
545                 550                 555                 560

Asp Ala Lys Glu Leu Ala Asn Ile Leu Lys Tyr His Ile Gly Asp Glu
                565                 570                 575

Ile Leu Val Ser Gly Gly Ile Gly Ala Leu Val Arg Leu Lys Ser Leu
            580                 585                 590

Gln Gly Asp Lys Leu Glu Val Ser Leu Lys Asn Asn Val Val Ser Val
        595                 600                 605

Asn Lys Glu Pro Val Ala Glu Pro Asp Ile Met Ala Thr Asn Gly Val
    610                 615                 620
```

FIG. 1C

```
Val His Val Ile Thr Asn Val Leu Gln Pro Pro Ala Asn Arg Pro Gln
625                 630                 635                 640

Glu Arg Gly Asp Glu Leu Ala Asp Ser Ala Leu Glu Ile Phe Lys Gln
                645                 650                 655

Ala Ser Ala Phe Ser Arg Ala Ser Gln Arg Ser Val Arg Leu Ala Pro
            660                 665                 670

Val Tyr Gln Lys Leu Leu Glu Arg Met Lys His
        675                 680
```

FIG. 1D

EXPRESS MAIL NO. EL255095431US

PATENT

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant	:	Jiangchun Xu
Filed	:	July 2, 1999
For	:	COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS OF COLON CANCER AND METHODS FOR THEIR USE
		Docket No. : 210121 471C1
		Date : July 2, 1999

Box Patent Application
Assistant Commissioner for Patents
Washington, D.C. 20231

DECLARATION

Sir:

I, Lawrence Teague, in accordance with 37 C.F.R. § 1.821(f) do hereby declare that, to the best of my knowledge, the content of the paper entitled "Sequence Listing" and the computer readable copy contained within the floppy disk are the same.

I declare further that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true and further that these statements are made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issuing thereon.

Dated this 2nd day of July, 1999.

Lawrence Teague

Lawrence Teague
Legal Assistant

6300 Columbia Center
Seattle, WA 98104-7092
(206) 622-4900
FAX (206) 682-6031

ljt\sequence new rule\210121\471c1 dec

SEQUENCE LISTING

<110> Xu, Jiangchun

<120> COMPOUNDS FOR IMMUNOTHERAPY AND
DIAGNOSIS OF COLON CANCER AND METHODS FOR THEIR USE

<130> 210121.471C1

<140> US
<141> 1999-07-02

<160> 120

<170> FastSEQ for Windows Version 3.0

<210> 1
<211> 458
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(458)
<223> n = A,T,C or G

<400> 1
ncaggtctgg cggcacctgt gcactcagcc gtcgatacac tggtcgattg ggacagggaa 60
gacgatgtgg ttttcaggga ggcccagaga tttggagaag cggatgaagt tctcctttag 120
ttccgaagtc agctccttgg ttctcccgta gagggtgatc ttgaagtact ccctgttttg 180
agaaactttc ttgaagaaca ccatagcatg ctggttgtag ttggtgctca ccactcggac 240
gaggtaactc gttaatccag ggtaactctt aatgttgccc agcgtgaact cgccgggctg 300
gcaacctgga acaaaagtcc tgatccagta gtcacacttc tttttcctaa acaggacgga 360
ggtgacattg tagctcttgt cttctttcag ctcatagatg gtggcataca tcttttgcgg 420
gtctttgtct tctctgagaa ttgcattccc tgccagga 458

<210> 2
<211> 423
<212> DNA
<213> Homo sapien

<400> 2
cagggtccat aggtgatccg caactctcga gcatttatat acaatagcaa atcatccagt 60
gtgttgtaca gtctataata ctccaacagt ctcccatctg tattcaatgg cgccacccaa 120
tacagtcctt tgtttggatg ctggggagag taatccctac cccaagcacc atatagataa 180
gaaaaccctc tccagttgag ctgaaccaca gacggtttgc tgatgttcac cacaccacca 240
tgaccacagc tccctggagt gggaggaggg tggacgacag gggtgttttg atctttagag 300
gcttcacact ctttcagctt ggtcttcaga gccacgattt ctcggcgaat ggcaaggaca 360
ttgtttttgt ctagtgtctc aagcttctct accaagagag tcatatttct tatctccacc 420
tcc 423

<210> 3
<211> 538

<212> DNA
<213> Homo sapien

<400> 3
ggtctgtcca atggcaacag gaccctcact ctaytcartg tcacaagraa tgayrcagsa 60
msctayraat gtgaaaycca gaacccagtg agtgccarsc gcagtgayyc agtcatcctg 120
aatgtcctct atggcccrga tgmccccacc atttcccctc taaacacatm ttaccgwyca 180
ggggaaaatc tgaacctctc ctgccacgca gcctctaacc cacctgcaca gtactcttgg 240
tttrtcaatg ggactttcca gcaatccacm caagagctct ttatccccaa catcactgtg 300
aataatagyg gatcctatac gtgccaagcc cataactcag mcactggcct caataggacc 360
acagtcacga cgatcacagt ctatgcaaga gccacccaaa cccttcatca ccagcaacaa 420
ctccaacccc gtggaggatg aggatgctgt agccttaacc tgtgaacctg agattcagaa 480
cacaacctac ctgtggtggg taaataatca gagcctcccg gtcagtccca ggctgcag 538

<210> 4
<211> 309
<212> DNA
<213> Homo sapien

<400> 4
tggtaascca aaaagatgct ggggcagatt gtggacaagt agaagaacct ccttcccctc 60
tgcgaacatt gaacggcgtg gattcaatag tgagcttggc agtggtgggc gggttccaga 120
aggttagaag tgaggctgtg agcaggagcc cctgccaggg gatvcacgca mtctgtgggg 180
aggggctgag rggdgwcycc atggtctctg ctgtctgctc tgtcctcctc tgtggagaag 240
agcttgagct ccaggaacgc tttgrtcavg gctgcctgtg acctytgctc tgbtctgcct 300
gcccgggcg 309

<210> 5
<211> 412
<212> DNA
<213> Homo sapien

<400> 5
gtccaatggc aacaggaccc ctcacttcta ttcaatgtca caagaaatga cgcaagagcc 60
tatgtatgtg gaatccagaa ctkcagtgag tgcaaaccgc agtgacccag tcaccctgga 120
tgtcctctat gggccagaca sccccccatca tttccccccc agactcgtct tacctttcgg 180
gagcgaacct caacctctcc tgccactcgg cctctaaccc atccccgcag tattcttggc 240
kgtatcaatg ggataccgca gcaacacaca caagttctct ttatcgccaa aatcacgcca 300
aataataacg ggacctatgc ctgttttgtc tctaacttgg ctactggccc gcaataattc 360
catagtcaag agcatcacag tcttctgcat ctggaacttc tcctggtctt ct 412

<210> 6
<211> 332
<212> DNA
<213> Homo sapien

<400> 6
gtgcaagggc tttacaaaaa ctgtgccagt krcttctyca tgwsrcwrga tctgacttka 60
ttsaygttkt atgagsysya saatmctgaw gctcmttyts sakgrwsttc kgsatmrgca 120
gtsrattcsa catttgggrt akrtymtctc tsgaagysam tgtcakgcag tgrcayccwr 180
gkktcwgcwt gcwgtgrgtt amcakcmwtr ywtagkgsgm ayatrattta ramrgtayak 240
cymtctcmct cytycmccay wtgcwcaass mkcacacctc ggccgcgacc acgctaagcc 300
cgaattccag cacactggcg gccgttacta gt 332

<210> 7

<211> 401
<212> DNA
<213> Homo sapien

<400> 7
tggtgttgtt ggcgccagtt ccctggacct ggaacagccg tgtggagggc ccggtctcca 60
agttgttagt tcgggaggtg cctccctggt agaccaccat gcgtcccttg aagatggaca 120
taagatgagg tggctccttg cccattggga cccggatctg gactggttca ccattgtact 180
tctggtccag gatgacggct tgataagctg atgctgtaat ttcatcttgg ctggcctggc 240
tgccctgcca aacgtagagc aggtaatgct gcttctcgcc gatgaaggta ggtgtaagag 300
cagcaggtaa gcaagttcgc ccccatagaa gtgggcctag ccacttggaa ttccagcaca 360
ctggcggccc gttactagtg ggatcccgag ctcggtacca a 401

<210> 8
<211> 1151
<212> DNA
<213> Homo sapien

<400> 8
ctctctccat aaaactcagc actttacaga tgtagaatat ataagcatgc caaatttact 60
tatctgccac atacaaagca tcattccagg tgctagtgag gggaaaaaaa agttggagat 120
ttggtccctc gaggagctcc agatattaat ctacctaact aagtccccag gtttcttcca 180
ggcatggaag aattagtggt gctacatgga tgaggactag tcattgggca atatttcctg 240
tacaaagaat ccctagacgc catactgagt tttaagttcc ttaattccta atttaaggct 300
tctagtgaag cctcctcaca gtaggcttca ctaggcccac agtgccccta gacctctgac 360
aatcccaccc tagacagact ttattgcaaa atgcgcctga agaggcagat gattcccaag 420
agaactcacc aaatcaagac aaatgtccta gatctctagt gtggtagaac tatgcaccta 480
aacattgctg caaaatgaac acacttttag acacccctgc agatatctaa gtaagtggag 540
aagactattt tttcaacaaa cattttctct ttcaccctaa ctcctaaaca gcttactggg 600
gcttctgcaa gacagaaaga tcataattca gaaggtaacc atcgttatag acataaagtt 660
tctggtcaaa agggttatag ttaatgctct gcacttttтc ctgcatctta tgcattacaa 720
tgtctagttt gccctctttc cctgtgtttg tgtcataata gtaaaaaatc tcttctgttc 780
tggtgtttca tagtacgggt ggcatacaga accccacata ccatgaaggc gttagaagca 840
gatggtttat actgcttggt ataccaagtg tttagcacct gaagtgtggt gtcattgagt 900
ttactaatca ccatgttacc agtgctggct tcagttgaat aaataaccca caatccattc 960
tcatccacag caaagtcaat atcttgccaa gcaacattag catatgaaaa gcggttatta 1020
taggcagcat tagggagagt ttgagtcaca gcaatcgtgt tggtggtcag gttaactctg 1080
gcaatattcc cggtgttgta catgttgacg tacatgttgt tgttgtaaac tgctgtacca 1140
ctaccttgga c 1151

<210> 9
<211> 604
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(604)
<223> n = A,T,C or G

<400> 9
ctgtgcaagg gctttacaaa aactgtgcca ggacttccca tgaggctgga ttgcttgatt 60
catgttttat gagccccaca atactgaagc tccttttcca gggacttggc ataggcagtc 120
aattccacat ttgggatagg tcctctctgg aagtgaatgt caggcagtga catccaagtt 180
tctgcatgca gtgggttaac agccatgttt agggggaaca tgatttaaaa agtacatctc 240

tctccctcct cccccacatg cacaaggctc acatctcatt atggtgkcgg cccatgtcac 300
attaaagtgt gatacttkgg ttttgaaaac attcaaacag tctctgtgga aatctggaga 360
gaaattggcg gagagctgcc gtggtgcatt cctcctgtag tgcttcaagn taatgcttca 420
tcctttntta ataacttttg atagacaggg gctagtcgca cagacctctg ggaagccctg 480
gaaaacgctg atgcttgttt gaagatctca agcgcagagt ctgcaagttc atcccctctt 540
tcctgaggtc tgttggctgg aggctgcaga acattggtga tgacatggac cacgccattt 600
gtgg 604

<210> 10
<211> 473
<212> DNA
<213> Homo sapien

<400> 10
tcgagaagat ccctagtgag actttgaacc gtatcctggg cgacccagaa gccctgagag 60
acctgctgaa caaccacatc ttgaagtcag ctatgtgtgc tgaagccatc gttgcggggc 120
tgtctgtgga gaccctggag ggcacgacac tggaggtggg ctgcagcggg gacatgctca 180
ctatcaacgg gaaggcgatc atctccaata aagacatcct agccaccaac ggggtgatcc 240
actacattga tgagctactc atcccagact cagccaagac actatttgaa ttggctgcag 300
agtctgatgt gtccacagcc attgaccttt tcagacaagc cggcctcggc aatcatctct 360
ctggaagtga gcggttgacc ctcctgggct cccctgaatt ctgtattcaa agatggaacc 420
cctccaattg atgcccatac aaggaatttg cttcggaacc acataattaa aga 473

<210> 11
<211> 411
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(411)
<223> n = A,T,C or G

<400> 11
tcctcattgg tcggggccaa aagcgtgtac tggccgttac cttcaagcat cgtgttgagc 60
cctgatgcag ccacagcagc ccgaagggtc tcaaaggtgt cctcgatctc aatgatctgc 120
tggatgttgt tggtgatggt ggagatgacc ttatcgatga ggtgcaccac cccgttggtt 180
gcatggtggt cggctttyar carccgggca cagttcacag ttacaatccc attaggatag 240
tggtggatct nggatgttgg aattctggta catagnaggt gaggggtcat gcccgtgttt 300
cagctcatca gtcaggactc gcctgcccac catatggtaa gcsgragggc atttgagcag 360
ctcaatgttt gacattgctg gaccagggga gttccagcac ttctangang a 411

<210> 12
<211> 560
<212> DNA
<213> Homo sapien

<400> 12
tacttgcctg gagatwgcyt tykckwtmtg ytcwrawgtc cgtggataca gaaatctctg 60
caggcaagtt gctccagagc atattgcagg acaagcctgt aacgaatagt taaattcacg 120
gcatctggat tcctaatcct tttccgaaat ggcaggtgtg agtgcctgta taaaatattc 180
tatgtttacc ttcaacttct tgttctggct atgtggtatc ttgatcctag cattagcaat 240
atgggtacga gtaagcaatg actctcaagc aatttttggt tctgaagatg taggctctag 300
ctcctacgtt gctgtggaca tattgattgc tgtaggtgcc atcatcatga ttctgggctt 360
cctgggatgc tgcggtgcta taaaagaaag tcgctgcatg cttctgttgt ttttcatagg 420

cttgcttctg atcctgctcc tgcaggtggg cgacaggtat cctaggagct gttttcaaat 480
ctaagtctga tcgcattgtg aatgaaactc tctatgaaaa cacaaagctt ttgagcgcca 540
caggggaaag tgaaaaacaa 560

<210> 13
<211> 150
<212> DNA
<213> Homo sapien

<400> 13
gggcaggctg tctttttaaa atgtctcggc tagctagacc acagatatct tctagacata 60
ttgaacacat ttaagatttg agggatataa gggaaaatga tatgaatgtg tatttttact 120
caaaataaaa gtaactgttt acgttggtga 150

<210> 14
<211> 403
<212> DNA
<213> Homo sapien

<400> 14
ctgctgcctg tggcgtgtgt gggctggatc ccttgaaggc tgagtttttg agggcagaaa 60
gctagctatg ggtagccagg tgttacaaag gtgctgctcc ttctccaacc cctacttggt 120
ttccctcacc ccaagcctca tgttcatacc agccagtggg ttcagcagaa cgcatgacac 180
cttatcacct ccctccttgg gtgagctctg aacaccagct ttggcccctc cacagtaagg 240
ctgctacatc aggggcaacc ctggctctat cattttcctt ttttgccaaa aggaccagta 300
gcataggtga gccctgagca ctaaaaggag gggtccctga agctttccca ctatagtgtg 360
gagttctgtc cctgaggtgg gtacagcagc cttggttcct ctg 403

<210> 15
<211> 688
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(688)
<223> n = A,T,C or G

<400> 15
caaagcacat tttaatcatt tattttaaaa gggggagtaa agcatttaaa ctgccaatcc 60
tatagactag gacttgaaca tcaaaggaaa aatagacaaa gactagatga taaagtcatt 120
caaaagcaca gaagcacatc acatacacca gcaaggtttc caactactgc actgattaac 180
tagatactct caatagcttt tctatagctc gtcctagaaa aaaaaattaa attttcattt 240
tcttacaagt tccaggctta aacaaaggca aaaattacat gcaacaactg atacactcat 300
aagttgcaca tatgctccaa ggtctttatt agataacaat aaatgctagc actttgtcac 360
tgccatcaga ttttccttat agtcttagag tcatgtaaat aaaagttcca taatgaaatt 420
aaagaaaatt aatttttcta atcttagatc agttccatag aaaactatta atttttttaa 480
agtaggcagt agaagggggt tggtgggggg tggaattggt tagtaagtct ggttctaatc 540
ttctgagctg cctttggaag gaagttatga ggtagaagat tctactgact tttagtaagg 600
tggacaatga gagaaaagaa aaagcaggtg cctcatcnnc agatccttnt ggtatttatn 660
tgccangtnc nanntaatnc atanaaag 688

<210> 16
<211> 408
<212> DNA

<213> Homo sapien

<400> 16

caggtcatca agatgactta caggatgtaa tagggagagc tgtcgagatt ggtgttaaaa 60
agtttatgat tacaggtgga aatctacaag acagtaaaga tgcactgcat ttggcacaaa 120
caaatggtat gtttttcagt acagttggat gtcgtcctac aagatgtggt gaatttgaaa 180
agaataaccc tgatctttac ttaaaggagt tgctaaatct tgctgaaaac aataaaggga 240
aagttgtggc aataggagaa tgcggacttg attttgaccc gactgcagtt ttgtcccaaa 300
gatactcaac tcaaatattt tgaaaaacag tttgaactgt cagaacaaac aaaattacca 360
atgtttcttc attgtccgaa actcacatgc tgaatttttg gacataat 408

<210> 17
<211> 407
<212> DNA
<213> Homo sapien

<400> 17

ggtcctgggg aggccctagg ggagcaccgt gatggagagg acagagcagg ggctccagca 60
ccttctttct ggactggcgt tcacctccct gctcagtgct tgggctccac gggcaggggt 120
cagagcactc cctaatttat gtgctatata aatatgtcag atgtacatag agatctattt 180
tttctaaaac attcccctyc ccactcctct cccacagagt gctggactgt tccaggccct 240
ccagtgggct gatgctggga cccttaggat ggggctccca gctcctttct cctgtgaatg 300
gaggcagaag acctccaata aagtgccttc tgggcttttt ctaacctttg tcttagctac 360
ctgtgtactg aaatttgggc ctttggatcg aatatggtca agaggtt 407

<210> 18
<211> 405
<212> DNA
<213> Homo sapien

<400> 18

tgaagagtca acttgggcct ggaggactga taaagtttgt gattttgagg gcctctaaaa 60
gtattaaagc agcggcagcc gctgcacgca gacatgaggg ctaggttaaa acagtaagat 120
caagttgttt ggacagaaag gctacagagt gtggtcctgg ctcttgtgta agaattacga 180
ccacgctaac catgcctagg aaggaaagga gttattgttt tgtagaaagg tgctggggtt 240
tgagagatca gtcggacacg attggcaggg agagcacgtg tgtttttatg agaattatgc 300
ccgagatagg taacagatga ggaagaaatt tgggcttgat tgaagtaatg ggggctgtct 360
gtgaagcttt gcagcagtac agcctaggta atttgctgag cctaa 405

<210> 19
<211> 401
<212> DNA
<213> Homo sapien

<400> 19

tcctgacatt cctgccttct tatattaata agacaaataa aacaaaatag tgttgaagtg 60
ttggggcagc gaaaattttt ggggggtggt atggagagat aatgggcgat gtttctcagg 120
gctgcttcaa gcgggattag gggcggcgtg ggagcctaga gtgggagaga ttaagctgaa 180
gggaggtctt gtggtaaggg gtgatatcat ggggatgtta gaagaaacat ttgtcgtata 240
gaatgattgg tgatggcctg gatacggttt tggatgattt gagaagctaa atggaagata 300
caaggtccga ataaaaggag gagaaaaatg ggtattaaat gtctaagaat tgggaggacc 360
taggacatct gattagagag tgcctaagga gattcagcat a 401

<210> 20
<211> 331

<212> DNA
<213> Homo sapien

<400> 20
aggtccagct ctgtctcata cttgactcta aagtcatcag cagcaagacg ggcattgtca 60
atctgcagaa cgatgcgggc attgtccaca gtatttgcga agatctgagc cctcaggtcc 120
tcgatgatct tgaagtaatg gctccagtct ctgacctggg gtcccttctt ctccaagtgc 180
tcccggattt tgctctccag cctccggttc tcggtctcca ggctcctcac tctgtccagg 240
taagaggcca ggcggtcgtt caggctttgc atggtctcct tctcgttctg gatgcctccc 300
attcctgcca gacccccggc tatcccggtg g 331

<210> 21
<211> 346
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(346)
<223> n = A,T,C or G

<400> 21
ggtccaccac ttgtacccga tatggacttc cggcttctct gtccaatgga gccacactaa 60
agatctcacc agtcacgtgg tcaattttaa gccaacctct tgtgtctccc ctcagtgaat 120
agcttatgtc cagaccttct ggatccttgg cagtcacatt gcccacttta gtgcctatag 180
ctacatcctc actgactttc gcttggaata cgtgttggga aaattgaggt gcttcattca 240
catctgtcac aataagncgt gaacttggca aaagaacttg cattgtactt cacaccaaac 300
actagaggct caggattttc tgctttgaac acaatgttgg aaacag 346

<210> 22
<211> 360
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(360)
<223> n = A,T,C or G

<400> 22
gaagactccc tctctcggaa gccggatccc gagccgggca ggatggatca ccaccagccg 60
gggactgggc gctaccaggt gcttcttaat gaagaggata actcagaatc atcggctata 120
gagcagccac ctacttcaaa cccagcaccc gcagattgtg caggctgcgt cttcagcacc 180
agcacttgaa actgactctt cccctccacc atatagtagt attactggtg gaagtaccta 240
caacttcaga tacagaagtt tacggtgagt tttatcccgt gccacctccc tatagcgttg 300
ctacctctct tcctacnwta cgatgaaagc tgagaaggct aaagctgctg caatggcatg 360

<210> 23
<211> 251
<212> DNA
<213> Homo sapien

<400> 23
ggcggagctc cacgacgagc tggaaaagga acctttgag gatggctttg caaatgggga 60
agaaagtact ccaaccagag atgctgtggt cacgtatact gcagaaagta aaggagtcgt 120

gaagtttggc tggatcaagg gtgtattagt acgttgtatg ttaaacattt ggggtgtgat 180
gcttttcatt agattgtcat ggattgtggg tcaagctgga ataggtctat cagtccttgt 240
aataatgatg g 251

<210> 24
<211> 421
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(421)
<223> n = A,T,C or G

<400> 24
caggtctttc ccaggtgttg actccagctc cagcttcagc tccagctcca ggtcgggctc 60
cagctccagc cgcagcttar gcagcgggag gttctgtgtc ccagttgttt tccaatttca 120
ccggctcccg tggatgamcg ygggacctgy caswgctcct gtktycctgc yagsacacca 180
cnytttyccg tggacacrar kggaacckct tggaattcac agctyatgtt ctttctcara 240
agtttgagaa agaactttct aaagtgaggg aatatgtcca attaattagt gtgtatgaaa 300
agaaactgtt aaacctaact gtccgaattg acatcatgga raaaggatac catttcttac 360
actgaactgg acttcgagct gatcaaggta gaagtgaagg agatggaaaa actggtcata 420
c 421

<210> 25
<211> 381
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(381)
<223> n = A,T,C or G

<400> 25
gaacttttg tttctttatt ttcaatattt gtcttattaa tatttttctt attttataat 60
gcaattacaa caatttagga nacaaaacaa tataaacaaa agaatgttaa atagtttttt 120
ttaaaaaata gcttgttgct tgcaanaaag tccatataat cttattcccc cccaaatata 180
attttatact ttgcactaaa ccaaaatagc ttatggaaaa ttagtattaa atagctaaac 240
acagaaaacc tacagctata aataacataa aatacagttt aactttaatg ngatgcttaa 300
acaaagcaaa ctatgatgca atatgaatca acttcattaa ttggacaagt ccagnggagg 360
cacaaattag ataagcacta a 381

<210> 26
<211> 401
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(401)
<223> n = A,T,C or G

<400> 26
ggaaaaggga ctggcctctc tgaagagtga gatgagggaa gtggaaggag agctggaaag 60

gaaggagctg gagtttgaca cgaatatgga tgcagtacag atggtgatta cagaagccca 120
gaaggttgat accagaagcc aagaacgctg gggttacaat ccaagacaca ctcaacacat 180
tagacgggct cctgcattct gatggaccaa ccttttcang tggtaagatt gaaganggag 240
cctgggctta cctgggaagc aaaaactttt cccganccaa ggaacccagg attcaaccan 300
gcnacttgcn ggccaaggaa ggcanaactn ggaanaaaag gcccccttaag caaaagggnc 360
accttcattt gctnggaaan cagcctttan ttggaatctt g 401

<210> 27
<211> 383
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(383)
<223> n = A,T,C or G

<400> 27
aattgcaact ggacttttat tgggcagtta cnacaacnaa tgttttcana aaaatatttg 60
gaaaaaatat accacttcat agctaagtct tacagagaan aggatttgct aataaaactt 120
aagttttgaa aattaagatg cnggtanagc ttctgaacta atgcccacag ctccaaggaa 180
nacatgtcct atttagttat tcaaatacca gttgagggca ttgtgattaa gcaaacaata 240
tatttgttan aactttgntt ttaaattact gntncttgac attacttata aaggagnctc 300
taactttcga tttctaaaac tatgtaatac aaaagtatan ntttccccat tttgataaaa 360
gggccnanga tactgantag gaa 383

<210> 28
<211> 401
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(401)
<223> n = A,T,C or G

<400> 28
ggtcgcgttt cccctggctc acagtctgcc attatttgca tttttaaatg aagaaaagtt 60
taacgtggat ggatggacag tttacaatcc agtggaagaa tacaggaggc agggcttgcc 120
caatcaccat tggagaataa cttttattaa taagtgctat gagctctgcg acacttaccc 180
tgctcttttg gtggttccgt atcgtgcctc anatgatgac ctccggagag ttgcaacttt 240
taggtcccga aatcgaattc cagtgctgtc atggattcat ccagaaaata agacggtcat 300
tgtgcgttgc agtcagcctc ttgtcggtat gagtgggaaa cgaaataaag atgatgagaa 360
atatctcgat gttatcaggg agactaataa acaaatttct a 401

<210> 29
<211> 401
<212> DNA
<213> Homo sapien

<400> 29
atatgagttt gccatctcca tggatgccat ttcaatgcct tcagggtaat cattctctcc 60
ccaaagactg cccacggggt catcactcct gtgacgaaat gagggctgga ttgaagatgt 120
tctgctgagc acccccctgg tcatctttgg ggtctcagaa gagccataat catgaccatt 180
ctcagcatct gaataatcag gttctctcca agtgcttggc aagttctgat tgtcctcagc 240

actgggatag tctggctccc caaaaaaggg tggagagtta ggttgaatgt cagcgcctgg 300
ataatcaggc tttcccagag agtctgcgta tggattgatt ctaaaacttg tatgttccag 360
attctttctg gatcctggat ggttcaaatt ggctctgggt c 401

<210> 30
<211> 401
<212> DNA
<213> Homo sapien

<400> 30
cctgaactat ttattaaaaa catgaccact cttggctatt gaagatgctg cctgtatttg 60
agagactgcc atacataata tatgacttcc tagggatctg aaatccataa actaagagaa 120
actgtgtata gcttacctga acaggaatcc ttactgatat ttatagaaca gttgatttcc 180
cccatcccca gtttatggat atgctgcttt aaacttggaa gggggagaca ggaagtttta 240
attgttctga ctaaacttag gagttgagct aggagtgcgt tcatggtttc ttcactaaca 300
gaggaattat gctttgcact acgtccctcc aagtgaagac agactgtttt agacagactt 360
tttaaaatgg tgccctacca ttgacacatg cagaaattgg t 401

<210> 31
<211> 297
<212> DNA
<213> Homo sapien

<400> 31
acctccatta atgccaggtg ttcctcctct gatgccagga atgccaccag ttatgccagg 60
catgccacct ggattgcatc atcagagaaa atacacccag tcattttgcg gtgaaaacat 120
aatgatgcca atgggtggaa tgatgccacc tggaccagga ataccacctc tgatgcctgg 180
aatgccacca ggtatgcccc cacctgttcc acgtcctgga attcctccaa tgactcaagc 240
acaggctgtt tcagcgccag gtattcttaa tagaccacct gcaccaacag caactgt 297

<210> 32
<211> 401
<212> DNA
<213> Homo sapien

<400> 32
caaacctgga gccaaaaagg acacaaagga ctctcgaccc aaactgcccc agaccctctc 60
cagaggttgg ggtgaccaac tcatctggac tcagacatat gaagaagctc tatataaatc 120
caagacaagc aacaaaccct tgatgattat tcatcacttg ggtgagtgcc cacacagtca 180
agctttaaag aaagtgtttg ctgaaaataa agaaatccag aaattggcag agcagtttgt 240
cctcctcaat ctggtttatg aaacaactga caaacacctt tctcctgatg gccagtatgt 300
ccccaggatt atgtttgttg acccatctct gacagttaga gcccgatatc actggaagat 360
attcaaaccg tctctatgct tacgaacctg cagatacagc t 401

<210> 33
<211> 401
<212> DNA
<213> Homo sapien

<400> 33
agcagaggga caggaatcat tcggccactg ttcagacggg agccacaccc ttctccaatc 60
caagcctggc cccagaagat cacaaagagc caaagaaact ggcaggtgtc cacgcgctcc 120
aggccagtga gttggttgtc acttactttt tctgtgggga agaaattcca taccggagga 180
tgctgaaggc tcagagcttg accctgggcc actttaaaga gcagctcagc aaaaagggaa 240
attataggta ttacttcaaa aaagcaagcg atgagtttgc ctgtggagcg gtgtttgagg 300

agatctggga ggatgagacg gtgctcccga tgtatgaagg ccggattctg ggcaaagtgg 360
agcggatcga ttgagccctg gggtctggct ttggtgaact g 401

<210> 34
<211> 401
<212> DNA
<213> Homo sapien

<400> 34
aacaatggct atgaaggcat tgtcgttgca atcgacccca atgtgccaga agatgaaaca 60
ctcattcaac aaataaagga catggtgacc caggcatctc tgtatctgtt tgaagctaca 120
ggaaagcgat tttatttcaa aaatgttgcc attttgattc ctgaaacatg gaagacaaag 180
gctgactatg tgagaccaaa acttgagacc tacaaaaatg ctgatgttct ggttgcttga 240
gtctactcct ccaggtaatg atgaacccta cactgagcag atggggcaac tgtggagaga 300
aggggtgaaa ggatcccacc tcactcctga tttcattgca ggaaaaaagt tagcttgaat 360
atggaccaca aggtaagggc atttgtccat gaatggggct c 401

<210> 35
<211> 401
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(401)
<223> n = A,T,C or G

<400> 35
catttcttcc tactagactg ccccccttgat ccactggcag aaatgatggc accaccttgt 60
cttcaggtgg tgctccttca ttattccaag gatgcagcat ctctatggtg ccaggtatgg 120
gggtaaagcc tttggcgccc tttccgcaat ggcacatcag cagtaaaagt ggtaccaata 180
gcangaacag aaagggcaaa atcatgancg caattgctgc gggtcccaag cccacatagg 240
aatcatgctg ngcttccctg canccgctgc catgcaagac actnacaaac tgngantgta 300
aggacctgct tttcaggaca actaaaaccc tgattgnctg aaatcaggaa ctgaatttca 360
cttctcccaa gctttttctc actttggtgc aacancacac t 401

<210> 36
<211> 401
<212> DNA
<213> Homo sapien

<400> 36
cctgctagaa tcactgccgc tgtgctttcg tggaaatgac agttccttgt ttttttgtt 60
tctgtttttg ttttacatta gtcattggac cacagccatt caggaactac cccctgcccc 120
acaaagaaat gaacagttgt agggagaccc agcagcacct ttcctccaca caccttcatt 180
ttgaagttcg ggtttttgtg ttaagttaat ctgtacattc tgtttgccat tgttacttgt 240
actatacatc tgtatatagt gtacggcaaa agagtattaa tccactatct ctagtgcttg 300
actttaaatc agtacagtac ctgtacctgc acggtcaccc gctccgtgtg tcgccctata 360
ttgagggctc aagctttccc ttgtttttttg aaaggggttt a 401

<210> 37
<211> 401
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(401)
<223> n = A,T,C or G

<400> 37
cnnctntgna atggantnnt tgnctaaaan ganttgatga tgatgaanat ccctangang 60
antaagcatg gancntgatc ntttnctnng cactccttta cgacacggaa acangnatca 120
ncatgatggt accaganacc ttatcaccna cgcgcacnga nctgactnat tccaaagagt 180
tgnggttacg gncatccggt cattgctcgt gcccattgct gcagggctga tnctactggt 240
gcttattatg ntggccctga ggatgctcca caatgaatat aagcatgctg catgatcagc 300
ggcaacanat gctctgccgt ttgcactaca tctttcacgg acacnatntc gaanacgggc 360
acnttgcana gttagacttg gaatgcatgg ngccggncan n 401

<210> 38
<211> 401
<212> DNA
<213> Homo sapien

<400> 38
aattggctca ctctctcaag gcaagcactg tctcaaggca gtctcaaggc agagatgaca 60
cagcaaaaaa cagaggggga gaaaaaagtc tattattggc ttgtgattta caaaagccaa 120
agtcctttag ataaaaggcc aggagtcgta ccaacataga taccaaatcc aggagaacac 180
agaccagcga taagagggac gcttccccat gacccagacc agcctaaagc ccctgtgggg 240
gcagccagtg gggagctgtc agaccttgga catggtggtc tttgagaatg ggtctgccct 300
tctctccctg accagttggg atagacacct gactggaatc cttgacactg gcaggtgttt 360
ctatgaacag agaggactgt gcctgtcttc ctgaatccca a 401

<210> 39
<211> 401
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(401)
<223> n = A,T,C or G

<400> 39
tctggtangg agcaattcta ttatttggca ttgcatggct gggttgaatt aaaacaggga 60
gtgagaacag gtgagtctag aagtccaact ctgaaaagga ccactgtaca tttgaacaca 120
cggctgtgtt aaagatgctg ctaatgtcag tcactgggtg cactaaagga tctcttattt 180
tatgtaaaac gttgggaatg acaagatana actgatactc tggtaagtta ccctctgaag 240
ctacttcttg tgaaatacta atgacagcat catcctgcca agcgaaagag gcaggcataa 300
gcaaggacaa attaaaaggg ggtaagagcc ttatcatgat gaggagtctt gttttgacat 360
cttgggaaaa gctgtccata gtgtgaagtc gtcaatttct c 401

<210> 40
<211> 401
<212> DNA
<213> Homo sapien

<400> 40
tctggtcacc caactcttgt ggaagagggg aattgagatc gagtactgaa tatctggcag 60
agaggctgga atccttcagc cccagagccc agggaccact ccagtagatg cagagagggg 120

cctgcccagg ggtcagggca gtgggtatca ctggtgacat caagaatatc agggctgggg 180
aggcatcttt gtttcctggt gccctcctca aagttgctga cactttgggg acgggaaggg 240
gtagaagtag ggctgctcct tttggagctg gagggaatag acctggagac agagttgagg 300
cagtcgggct gtccaggttc taagcatcac agcttctgca ctgggctctg aggagattct 360
cagccagagg atcccagcct cctcctccct caaatgtcaa g 401

<210> 41
<211> 401
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(401)
<223> n = A,T,C or G

<400> 41
ctggactaaa aatgtccact atggggtgca ctctacagtt tttgaaatgc taggaggcag 60
aaggggcaga gagtaaaaaa catgacctgg tagaaggaag agaggcaaag gaaactaggt 120
ggggaggatc aattagagag gaggcacctg ggatccacct tcttccttan gtcccctcct 180
ccatcagcaa aggagcactt ctctaatcat gccctcccga agactggctg ggagaaggtt 240
taaaaacaaa aaatccagga gtaagagcct taggtcagtt tgaaattgga gacaaactgt 300
ctggcaaagg gtgcganagg gagcttgtgc tcangagtcc agcccgtcca gcctcggggt 360
gtangtttct gaagtgtgcc attggggcct caccttctct g 401

<210> 42
<211> 310
<212> DNA
<213> Homo sapien

<400> 42
ggttcgacaa atccccaaaa atggcaaatt aagccctgtg acaaaataag ttattggatc 60
atacagaaat agcccaaatc tggaaatttt gaattaaaat tgtaatcctg taaaacaagt 120
tttggggtga atggatttct ttaataccaa taatattttt aattcccacc acagatggat 180
ttgctgaata tgctaatgct gtgaatgaga aaacaatttt ggggtaggta tacccacaag 240
taatctgatg acaaaataaa ccacagactg atgtcaaatg gacaaaaaac tgaaaatatg 300
ctgtgagaaa 310

<210> 43
<211> 401
<212> DNA
<213> Homo sapien

<400> 43
aggtcactta cacttgtgac cagtgtgggg cagagaccta ccagccgatc cagtctccca 60
ctttcatgcc tctgatcatg tgcccaagcc aggagtgcca aaccaaccgc tcaggagggc 120
ggctgtatct gcagacacgg ggctccagat tcatcaaatt ccaggagatg aagatgcaag 180
aacatagtga tcaggtgcct gtgggaaata tccctcgtag tatcacggtg ctggtagaag 240
gagagaacac aaggattgcc cagcctggag accacgtcag cgtcactggt attttcttgc 300
caatcctgcg cactgggttc cgacaggtgg tacagggttt actctcagaa acctacctgg 360
aagcccatcg gattgtgaag atgaacaaga gtgaggatga t 401

<210> 44
<211> 401
<212> DNA

<213> Homo sapien

<400> 44
atccctgtaa gtctattaaa tgtaaataat acatacttta caacttctct tagtcggccc 60
ttggcagatt aaatctttgc aaaattccat atgtgctatt gaaaaatgaa ataaaacctc 120
agatgtctga attcttattt caaatacagt tatataatta ttttaaatta caatatacaa 180
tttctgttaa atacaactgt taagggattc tgagaacaat tataagatta taataatata 240
tacaaactaa cttctgaaat gacatgggtt gtttccttcc caccctccta ccctctcaaa 300
gagtttttgc atttgctgtt cctggttgca aaaggcaaaa gaaaatctaa aaatagtctg 360
tgtgtgtcca cgacatgctc gctcctttga gaatctcaaa c 401

<210> 45
<211> 401
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(401)
<223> n = A,T,C or G

<400> 45
gtgcctgctg cctggcagcc tggccctgcc gctgcctcag gaggcgggag gcatgagtga 60
gctacagtgg gaacaggctc aggactatct caagagattt tatctctatg actcagaaac 120
aaaaaatgcc aacagtttag aagccaaact caaggagatg caaaaaattc tttggcctac 180
ctatactgga atggtaaact cccgcgtcat anaaataatg caanaagccc agatgtggag 240
tgccagatgt tgcagaatac tcactatttc caaatagccc aaaatggact tccaaagtgg 300
tcacctacag gatcgtatca tatactcgag acttaccgca tattacagtg gatcgattag 360
tgtcaaaggc tttaaacatg tggggcaaag agatccccct g 401

<210> 46
<211> 401
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(401)
<223> n = A,T,C or G

<400> 46
gtcagaattg tctttctgaa aggaagcact cggaatcctt ccgaactttc caagtccatc 60
catgattcan agatactgcc ttctctctct ctgggatttt atgtgtttct gatagtgaat 120
tgttgatgta tttgctactt tgcttctttt ctctttcaag acttgatcat tttatatgct 180
gnttggagaa aaaaagaact tttggtagca aggaggtttc aagaaatgat tttggatttt 240
ctgctgcgga atttctcggc acctacctgt agtatggggc acttggtttg gttgcagagt 300
aagaaggtgg aagaatgagc tgtacttggt taagcagttg aaacctttt tgagcaggat 360
ctgtaaaagc ataattgaat ttgtttcacc cccgtggatt c 401

<210> 47
<211> 401
<212> DNA
<213> Homo sapien

<400> 47

ggtctgcagc aatgcacttc aaccatacat actgcttcca ctagctaata ccaaatgcag 60
gttctcagat ccagacaaat ggaggaaaag aacatttatg cttccgtttc agaaagccaa 120
gtcgtagttt tggcccttcc tttctctaaa gtttattccc aaaaacaggt agcattcctg 180
attgggcaga gaagaggata ttttcagccc acatctgctg caggtatgtc attttctccc 240
atcttcactg tgactagtaa agatctcacc acttctcttt ggaatttcca actttgcttg 300
tgattgaatg tcacttcgtg aatttgtatt atgtcagatc acttggcatt gctcttccat 360
atgcatcaag ttgccaggca ctaaacccaa tgttcatgaa c 401

<210> 48
<211> 430
<212> DNA
<213> Homo sapien

<400> 48
acataacttg taaacttttt ctgcttgggg gctgtaacag acagaagagt aaagactaca 60
aggattttct gaagatgctt caatgaaaat catcatttcc tctttagtca tcccaagtct 120
tggtttgaaa aacttgggca tggacttata cagaccttga accaccactg acttatcatt 180
gggtggcaga ccttgaaacc aagctctctg tgttacttct gaaagtgcat caattctgat 240
ttggctaaga acagaagaca aatactggga tcgtgattct gtgttatact ctagccacag 300
catagcagct tctcgaacgg tttcttcctt ttctacattt aaattgtcac tactgagaat 360
atctatcagt aggtcatgtg acagacctgc cccggggccg gcccgctcga tgcttgccga 420
atatcatggt 430

<210> 49
<211> 57
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(57)
<223> n = A,T,C or G

<400> 49
ggtattaaca atatcangca ctcattcttc ccctcttatg aaanggatna atttta 57

<210> 50
<211> 327
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(327)
<223> n = A,T,C or G

<400> 50
gatggnggtn tccacaagan tnaangtncn tattaantan nncttgtaga nccacttnna 60
ttaattgnnn tatgnntgnc cttctggtgg ntgtngaagc ttcatatnnt ntttggacat 120
cattacacgt cttagctctt tnaagnacaa ctttaatgct atatgaattt tgccattttn 180
gctaacactg gtatgctccn ngcatccacc atnccacntg gaattattta ttncnttcat 240
attaatnttt tgtttaccaa atctnacttg acccgaacga aactttctgn gtattttang 300
gccccnccat tcttactttt caagcct 327

<210> 51

<211> 236
<212> DNA
<213> Homo sapien

<400> 51
cgtctcgaag aagcgctgca ggccgatgat ggactgcacg tctgccttgt cctcagttaa 60
cttgttgaat tgcttgaaca tgcggcccac atcctgggca aactcctgtg gggagctgta 120
gggaggtgac aacttctcct ggaggcgggc acggatcagg gtcagatcca gggtgccacc 180
gggctggtcc agggagaagg tggagtcgta gccagacctg cccgggcggc cgctcg 236

<210> 52
<211> 291
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(291)
<223> n = A,T,C or G

<400> 52
ctcacatcct gggtccggct gtagagctgc accatggtgc tgagcgcccc ctccagctcc 60
ttgtagatgt aaaggacggc gaaggagctg tagtctgtgt ccacgatgcg cacgtccagg 120
tagcccaagg ccgggactct gaagttgtcc ctcggagccc accttcangt actcgggcat 180
ccacctggtt acagccnttc gncctcggna actccatntg gactttacag gccgccctcc 240
tctgtgggcc tgatggncct tgcaggacat nggaacacgg gagctcnctt t 291

<210> 53
<211> 95
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(95)
<223> n = A,T,C or G

<400> 53
gtctgtgcag tttctgacac ttgttgttga acatggntaa atacaatggg tatcgctgan 60
cactaagttg tanaanttaa caaatgtgct gnttg 95

<210> 54
<211> 66
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(66)
<223> n = A,T,C or G

<400> 54
cctnaatnat ntnaatggta tcaatnnccc tgaangangg gancggngga agccggnttt 60
gtccgg 66

<210> 55
<211> 265
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(265)
<223> n = A,T,C or G

<400> 55
atctttcttc tcagtgcctt ggccntgttg agtctatctg gtaacactgg agctgactcc 60
ctgggaagag aggccaaatg ttacaatgaa cttaatggat gcaccaagat atatgaccct 120
gtctgtggga ctgatggaaa tacttatccc aatgaatgcc gtgttatgtt tttgaaaatc 180
ggaaacgcca gacttctatc ctcattcaaa aatctgggcc ttnctgaaaa ccagggtttt 240
naaaatccca ttcnggtcnc cggcg 265

<210> 56
<211> 420
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(420)
<223> n = A,T,C or G

<400> 56
gagcggccgc ccgggcaggt cctcgcggtg acctgatggg atttcaaaac cttggttctc 60
agcaaggccc agatttttga atgangatag aagtctggcg tttccgattt tcaaaacata 120
acacgcattc attgggataa gtatttccat cagtcccaca gacngggtca tatatcttgg 180
gtgcatccat taagttcntt tgttaacatt tgggcctctc tttcccangg gaattcagct 240
cccagttgtt taccaanatt naactccacc ggggccaaag gcncttgaaa aaaaaaanaa 300
ttccttgttt accttccttg ggcttnaagt tctggcgtcc aaaagttcaa tttgaaaact 360
gcaccgcact taccacgtct cttcnagaan cctggggaca cctcggccgc gaccacgcta 420

<210> 57
<211> 170
<212> DNA
<213> Homo sapien

<400> 57
gaagcggagt tgcagcgcct ggtggccgcc gagcagcaga aggcgcagtt tactgcacag 60
gtgcatcact tcatggagtt atgttgggat aaatgtgtgg agaagccagg gaatcgccta 120
gactctcgca ctgaaaattg tctctccaga cctcggccgc gaccacgcta 170

<210> 58
<211> 193
<212> DNA
<213> Homo sapien

<400> 58
attttcagtg cgagagtcta ggcgattccc tggcttctcc acacatttat cccaacataa 60
ctccatgaag tgatgcacct gtgcagtaaa ctgcgccttc tgctgctcgg cggccaccag 120
gcgctgcaac tccgcttcat cggcttcgcc cagctccgcc attgttcgcc acctgcccgg 180

gcggccgctc gaa 193

<210> 59
<211> 229
<212> DNA
<213> Homo sapien

<400> 59
cgcaactctc gagcatttat atacaatagc aaatcatcca gtgtgttgta cagtctataa 60
tactccaaca gtctcccatc tgtattcaat ggcgccaccc aatacagtcc tttgtttgga 120
tgctggggag agtaatccct accccaagca ccatatagat aagaaaaccc tctccagttg 180
agctgaacca cagacggttt gctgatacct gcccgggcgg ccgctcgaa 229

<210> 60
<211> 340
<212> DNA
<213> Homo sapien

<400> 60
tcgagcggcc gcccgggcag gtcctctaaa gatcaaaaca cccctgtcgt ccaccctcct 60
cccactccag ggaagctgtg gtcatggtgg tgtggtgaac atcagcaaac cgtctgtggt 120
tcagctcaac tggagagggt tttcttatct atatggtgct tggggtaggg attactctcc 180
ccagcatcca aacaaaggac tgtattgggt ggcgccattg aatacagatg ggaaactgtt 240
ggagtattat aaactggtac aacacactgg atgatttgct attgtatata aatgctcgag 300
aattgcggat cacctatgga cctcggccgc gaccacgctg 340

<210> 61
<211> 179
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(179)
<223> n = A,T,C or G

<400> 61
tttttgtgac ggacgnttgg agtacatgtc ccaggatcac atccagcagc tagagtggct 60
gggacaagct ggcggnggcc aagcactgtt gaaacnatag gggtctgggn gnactcgggt 120
tnaagtggtt ggtccgantn ttnataacct tgtcngaacc nancatctcg gttgncang 179

<210> 62
<211> 78
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(78)
<223> n = A,T,C or G

<400> 62
agggcgttcg taacgggaat gccgaagcgt gggaaaaagg gagcggtggc nggaagacgg 60
ggatgagctt angacaga 78

<210> 63
<211> 410
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(410)
<223> n = A,T,C or G

<400> 63
cccagttact tggggaggct gaggcaggga gaatcctttg aacccggngg gtgggaggtt 60
gcagtgagcc cgagatagca ccattgcact tccancatgg ggtggacaga gtgagactct 120
atctcaaaaa aaaagaaaag aaaaggaaag agattagatt aagattaagt acctacttcc 180
tntcccattt caagtcctga aaatagagga tcagaaatgt tgaggaattc tttaggatag 240
aaagggagat gggattttac ttatggggaa agaccgcaaa taaagactgn aacttaacca 300
cattccccaa gtgnaaggtg ttacccaaga agtaggaacc cttttggctn ttaccttacc 360
ttccngaaaa aaacttattn cttaaaatgg aaaacccttaa agcccgggca 410

<210> 64
<211> 199
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(199)
<223> n = A,T,C or G

<400> 64
cttgttctca aaaaggtcaa agggagcccg acgaggaata aatagcaatg ccctgaattc 60
caactgacct tctacagaaa agtgcttgac tgccaagtgg tcttcccagt cattagtgag 120
gctcttgtag aattctccat actcctcttg ggngangnca tnagggtttn nggcccaaat 180
aggntgggcc tngttaagt 199

<210> 65
<211> 125
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(125)
<223> n = A,T,C or G

<400> 65
agcggtacag ttctgtcctg gcatcatcat tcattgtagt atggtcaata ggtgccatga 60
aactcagtag cttgctaagg acatgaaacc gaagtttcct gcctttgctg gcctngtngn 120
gggta 125

<210> 66
<211> 204
<212> DNA
<213> Homo sapien

<400> 66
attcagaatt ctggcatcgg tatttctata aagtccatca gttagagcag gagcaggccc 60
ggagggacgc cctgaagcag cgggcggaac agagcatctc tgaagagccc ggctgggagg 120
aggaggaaga ggagctcatg ggcatttcac ccatatctcc aaaagaggca aaggttcctg 180
tggacctcgg ccgcgaccac gcta 204

<210> 67
<211> 383
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(383)
<223> n = A,T,C or G

<400> 67
tcagggcctc caggcagcca gttttgcagg anattcagca cctagngtct tcctgcctna 60
cgctcccaag aacctgctcc tgcaggggga acatcagaac tcgtccttga tgtcaaaatg 120
gggctggtct tnaggcttga agtccaggtt agggctgcca tcctcattga gaattctccg 180
ggcagtgtan ccgacgatgg ggtatttggc tttgtacact ttggtgaaaa cctnatccag 240
ggcctccagt tccttggccg tganacccgt antgtcatgg gtgaggtctg caggatccaa 300
ggacatcttg gctacccctc tagtggagtc cttccccgtc aaggcattgt aaggggctcc 360
tcgtccataa aactcctttt cgg 383

<210> 68
<211> 99
<212> DNA
<213> Homo sapien

<400> 68
tcacatctcc ttttttttttt aacttttttca aatttttgtg ttaaatagaa ggctaaaggg 60
ttagatttaa gtttctgcta cattgaccct atttaccta 99

<210> 69
<211> 37
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(37)
<223> n = A,T,C or G

<400> 69
gagaaggacn tacggncctg ntantanang aatctcc 37

<210> 70
<211> 222
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(222)

<223> n = A,T,C or G

<400> 70
gtgggtcatt tttgctgtca ccagcaacgt tgccacgacg aacatccttg acagacacat 60
tcttgacatt gaagcccaca ttgtccccag gaagagcttc actcaaagct tcatggcgca 120
tttcgacaga ttttacttcc gttgtaacgt tgactggagc aaaggtgacc accataccgg 180
gtttgagaac acccantcac ctgccccggg cggccgctcg aa 222

<210> 71
<211> 428
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(428)
<223> n = A,T,C or G

<400> 71
caggagtatt ttgtagaaaa gccagaagag cattagtaga tgtatggaaa tatacggtag 60
ggcacacgct gacagtactt ttcccaagcc acgccgtatt tcttcttaca gtggtactcg 120
tcacgagctt ctcggtggac aagcaacatg gtgaaataaa ttatgtagaa ataaggcaga 180
atgtggttaa aaccacatgg gagggaccac gccaaggcca tgatgagatc acccaagtaa 240
ttggggtggc gaacaaagcc ccaccatcca gaaactagaa naatttttcc cgttgaaata 300
tgaatggntt ttaaatgtgc aagctttgga tcactgggaa ttttcccgaa tgccttttc 360
tganaattgc accttnggaa gantccttac cccaagnttc agaccattat ttnaaaagcn 420
ttggaact 428

<210> 72
<211> 264
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(264)
<223> n = A,T,C or G

<400> 72
gaataaagag cttactggaa tccagcaggg ttttctgccc aaggatttgc aagctgaagc 60
tctctgcaaa cttgatagga gagtaaaaag ccacaataga gcagtttatg aagatcttgg 120
aggagattga cacacttgat cctgccagaa aatttcaaag acagtagatt gaaaaggaaa 180
ggctttggta aaaaaaggtt caggcattcc tagccgantg tgacacagtg gagcanaaca 240
tctgcangag actgancggc tgca 264

<210> 73
<211> 442
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(442)
<223> n = A,T,C or G

<400> 73
ggcgaatccg gcgggtatca gagccatcag aaccgccacc atgacggtgg gcaagagcag 60
caagatgctg cagcatattg attacaggat gaggtgcatc ctgcaggacg gccggatctt 120
cattggcacc ttcaaggctt ttgacaagca catgaatttg atcctctgtg actgtgatga 180
gttcagaaag atcaagccaa agaacttcaa acaagcagaa agggaagaga agcgagtcct 240
cggtctggng ctgctgccaa gggagaatct ggtctcaatg acngtagaag gaccttcttc 300
caaagatact ggnattgctc gagttccact tgctggaact tcccggggcc caaggatcgc 360
aaggcttctg gcaaaagaaa tccanacttn ggccgggacc acctaancca attcacacac 420
tggcggccgt actagtggat cc 442

<210> 74
<211> 337
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(337)
<223> n = A,T,C or G

<400> 74
ggtagcagcg tctccagagc ctgatctggg gtcccagata cccaggcagc agcagccctg 60
gaggtaaagg gcaagctccc caatgtgagg ggagacccca ttcctggtca gccaggcttt 120
cagaggagat agcaggtcga gggagccaac gaagaagaga ctgccancag gggaaggact 180
gtcccgccaa ggacagaact gattcagggg ggtcaatgct cctctagaga agagccacac 240
agaactgggg ggtccaggaa ccatgaanct tggctgtggt ctaaggagcc aggaatctgg 300
acagtgttct gggtcatacc aggattctgg aattgta 337

<210> 75
<211> 588
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(588)
<223> n = A,T,C or G

<400> 75
catgatgagt tctgagctac ggaggaaccc tcatttcctc aaaagtaatt tatttttaca 60
gcttctggtt tcacatgaaa ttgtttgcgc tactgagact gttactacaa actttttaag 120
acatgaaaag gcgtaatgaa aaccatcccg tccccattcc tcctcctctc tgagggactg 180
gagggaagcc gtgcttctga ggaacaactc taattagtac acttgtgttt gtagatttac 240
actttgtatt atgtattaac atggcgtgtt tatttttgta tttttctctg gttgggagta 300
tgatatgaag gatcaagatc ctcaactcac acatgtagac aaacattagc tctttactct 360
ttctcaaccc cttttatgat tttaataatt ctcacttaac taatttttgta agcctgagat 420
caataagaaa tgttcaggag agangaaaga aaaaaaatat atgttcccca tttatattta 480
gagagagacc cttantcttg cctgcaaaaa gtccaccttt catagtagta ngggccacat 540
attacattca gttgctatag gncagcactg aactgcatta cctgggca 588

<210> 76
<211> 196
<212> DNA
<213> Homo sapien

<400> 76
gcggtatcac agcctggccc ccatgtacta tcgggggccc caggctgcca tcgtggtcta 60
tgacatcacc aacacagata catttgcacg ggccaagaac tgggtgaagg agctacagag 120
gcaggccagc cccaacatcg tcattgcact cgcgggtaac aaggcagacc tggacctgcc 180
cgggcggccg ctcgaa 196

<210> 77
<211> 458
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(458)
<223> n = A,T,C or G

<400> 77
agtagagatg gggtttcact gtgttaacca ggatggtctt gatctcctgg cctcgtgatc 60
tgcccgcctc ggcctcccaa agtgttggga ttacaggcgt gaaccaccgc acccggccag 120
aaatgttagt ttttccctat tctctctcct ttttcctatt atatacttgg tcaaccagac 180
agccatccta ccccanaatg gtaatgcctc ttcattcctc atatgaggga ataaaagaga 240
aaaaagcttt tggaaaacat ccacttatct aatcatccca aatatgtaat caaaagtata 300
caactcatgt gaagaataca ctggtaaaat gttantatag gccaaggtat cttgaattcc 360
tatatagaaa gctggtaaat gccctttttgg ctggaaccgc catcttccnn taattcnccc 420
aaaatgacca aacacaaagg gnaagangan aagccccc 458

<210> 78
<211> 464
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(464)
<223> n = A,T,C or G

<400> 78
tccgcaaatt tcctgccggc aaggtcccag catttgaggg tgatgatgga ttctgtgtgt 60
ttgagagcaa cgccattgcc tactatgtga gcaatgagga gctgcgggga agtactccag 120
aggcagcagc ccaggtggtg cagtgggtga gctttgctga ttccgatata gtgcccccag 180
ccagtacctg ggtgttcccc accttgggca tcatgcacca caacaaacag gccactgaga 240
atgcaaagga ggaagtgagg cgaattctgg ggctgctgga tgcttacttg aagacgagga 300
cttttctggt gggcgaacga gtgacattgg ctgacatcac agttgtctgc accctgttgt 360
ggctctataa gcaggntcta gaaccttctt ttcgcangac cttcggccgg accacgctta 420
acccaaattc cacacacttg cnggccgtac taanggaatc ccac 464

<210> 79
<211> 380
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(380)
<223> n = A,T,C or G

<400> 79
ctgtatgacc agtttttcca tctccttcac ttctaccttg atcagctcga agtccagttc 60
agtgtaagaa atggtatcct tctccatgat gtcaattcgg acagttaggt ttaacagttt 120
cttttcatac acactaatta attggacata ttccctcact ttanaaagtt ctttctcaaa 180
cttctganaa aagaacatga actgtgaatt ccaagcgttc ccactctgtc cacgggaaaa 240
ggtggtgtct ggcagggaaa cagaacactg gcaggtccac ggtcatccac ggagccggtg 300
aaattgggaa aacaactggg acacagaacc tccgctgcct aagctgcggn tgggagcttg 360
gaacccgacc tggaactgga 380

<210> 80
<211> 360
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(360)
<223> n = A,T,C or G

<400> 80
tcgagcggcc gcccgggcag gtcctcagag agctgtttgt tncgcttctt caaaaactcc 60
tattctccac ttctgctaaa ggactggatg acatcaattg tgatagcaat atttgtgggt 120
gttctgtcan ncancatcgc actcctgaac aaagtagatg ttggattgga tcagtctctt 180
tccacccaga tgactcctan atggtggatn atttcaaatc catcantcag tacctgcatg 240
cgnggtccgc ctgtgtnctt tgtcctgcag gangggcnct actacacttc ttccnagggg 300
canaacatgg tgtgcngcgg ccatgggctg gcaacantga ttcnctgctg cacccanatn 360

<210> 81
<211> 440
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(440)
<223> n = A,T,C or G

<400> 81
acgtggtccg gcgagtctga cctgcagata tgaactcctt gggaaaccta cattctgcct 60
cagacatact gggggcaaat ggctttaaaa gtctggctca gggagccaag attacagaaa 120
nccgttgagt cnccatacat ggacactgac aaaggaactg aagatatcca aacaagccct 180
cctggtcccg ngcctgcata aagatcggga ncggaacggt accngacgtc tgtggtcagg 240
ggttgtggaa aattggaaaa aaccagtcct gcccacattg acagggaagc ctcaacggaa 300
attgaacaga tngtcttatc accagtctcc cctcctggat cntgtctcgg ctcnggggan 360
tcagtgatca gtcctttcag gtggaagaag caaagaagat caacaanaag cngatcctct 420
cacctgntac cagcatatgg 440

<210> 82
<211> 264
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature

<222> (1)...(264)
<223> n = A,T,C or G

<400> 82
agcgtggtcg cggccgangt cctgacattc ctgccttctt atattaatta tacnaataaa 60
acaaaatagt gttgaagtgt tggagcggcg aaaatttttg gggggtggta tggacagaga 120
atgggcgatn ttctcanggc tgcttcaagt gggattgggg cngcgtggga tcatncagtg 180
gganagattn cnctgaccgg antctnttgg tanggatnat cttgtgggga tgtgcaagag 240
ncattcgtct cctgaatgan tggt 264

<210> 83
<211> 410
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(410)
<223> n = A,T,C or G

<400> 83
ancgtggtcg cggccgangt ccacagttgt gggagagcca gccattgtgg gggcagctcc 60
acaggtaaga ctcgtgtcct gagcagcgca catcatccag gacaatgggt cctgagccct 120
gaccaaaccg ggcatttcct ggggctgaca tggcccagcc acagcccant tgcctgcaga 180
cgaaattggc atcattggtg tcccagtant catcacacac ggtgccccag gaacctccgg 240
tatangaact ccactcggcc tcnanacctg tcgcctccat tccncagcct caggggggcaa 300
actgggattc agatccttct gtgggtacag gtggtgatat cctgacaggc caactttctg 360
gcctgagtgt tgactgangc tgggcagacc tgcccgggcg gccgctcgaa 410

<210> 84
<211> 320
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(320)
<223> n = A,T,C or G

<400> 84
tcgaacggcc gcccgggcag gtctgcccca ggtgtatcca tttgccgccg atctctatca 60
naaggagctg gctaccctgc nncgacgaan tcctgaanat aatctcaccc ncccagatct 120
ctctgtcgca atggagatgt cgtcatcggt ggncctgatc acagggcatt ggactcagag 180
anangtnanc acagtgtnga agcgattgan nnagttcagt tgctggtctt acccgatntt 240
ggaaggaagg aaaacgtgtt angacgtatc tcgatgnant tgaccaaanc tgaangctnc 300
agggggcatc gcaaaganan 320

<210> 85
<211> 218
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(218)

<223> n = A,T,C or G

<400> 85
tcgagcggcc gcccgggcag gtctgctgcc cgtgctggtg ccattgcccc atgtgaagtc 60
actgtgccag cccagaacac tggtctcggg cccgagaaga ctcctttctc caggctntan 120
gtatcaccac taaaatctcc aggggcacca tnganatcct gggtgtccgc aatgttgcca 180
atgtctgtcc gcnnattggc tacccaactg ttgcatca 218

<210> 86
<211> 283
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(283)
<223> n = A,T,C or G

<400> 86
tcgacttctt gtgaaggttt tgganaaata tgtatcagtt cgttttattt gggtattcaa 60
taatatcctt ggtgataatg ctgactccat ggcttctgac cccaaaaatt gaccctgctg 120
ccactggttg tagccctgag attgattttt gtagccacga ttgtttcctc gtcctctgaa 180
gtnctggttg tanttccctc tgtngggcat tcccctctgt tgtanttccc tctgtttgan 240
taactaccac ggccaggaaa aacaggggca cgaaggtatg gat 283

<210> 87
<211> 179
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(179)
<223> n = A,T,C or G

<400> 87
agcgtggtcc cggccgatgt ctttctgtgt aagtgcataa cactccacat acttgacatc 60
cttcangtca cgggccagct nttcagcant ctctggagtg ataggctact gtntgttctn 120
ggcaagtgtc tcaanaatac aggggtcntc tctgagatga ntttcagtcc cgaaccctc 179

<210> 88
<211> 512
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(512)
<223> n = A,T,C or G

<400> 88
tcgagcggcc gcccgggcag gtcctancan agaatcacca aatttatgga gagttaacag 60
gggtttaaca ggaangaagt gcctttagta agttctcaag ccagangctg gaggcagcag 120
ctaaatcaga ggacaggatc ctcagtgaaa gtgagccatt cggggtggca tgtcactcca 180
ggaataagca caacttanaa acaaatgatt tcgtangata gcacagtgac attggtgcac 240

ttgtgaacct gaggccactg tgtcaaactg tgcactggtt gtgaataggg aganccaaaa 300
attatgtcct actgggtaat gagctttcaa tgggctcgat cctctcacnc tgaaagctct 360
gtagagcagc tcagaaccac aaccactccc aacattgacc cttctggggg tactgtctgt 420
ggcacccaca ggaaggagct ggagatcccc attaggactg tccacccaca cttgaagcca 480
caaaactgca cctcggccgc gaccaccgct ta 512

<210> 89
<211> 358
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(358)
<223> n = A,T,C or G

<400> 89
tcgagcgggc cgcccgggca ggtctgccag tccccatccc agacattctt tgcatctaag 60
ctgangtctg aactgagtgg ggtgggctgg tgtttccatc ctcacaactc cagtgagccg 120
ggtgtggccg tggcctgcgt ctctctggcg gttagtgatg ttggcatcat ccacctttt 180
caaaacaaaa gcactggact gaagaanaat cccnccctgt ntccacccag tccatggttt 240
ttaataaaag ggttatnnaa gttgancaag ncatcaccac acacaancct aagaacnttt 300
ttcatcnntc cccaaaacaa acccncaccc tgggaactcc gggcgcgaac cacgccta 358

<210> 90
<211> 250
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(250)
<223> n = A,T,C or G

<400> 90
cgagcggccg cccgggcagg tctggatggg gagacggact ggaactgcgg cttcccgtgg 60
cctgcacgca caaggctccc cacggccgcc gaccttcttc agattcgatc gtatgtgtac 120
gcacnaagag ccaaatattg acattcacaa cttcgtggga atnttacccc anaagactgc 180
gacccccccga tcaggcgana gcctgagcat agaagaacac cgctgtgggc ttggcactgt 240
gggncccatc 250

<210> 91
<211> 133
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(133)
<223> n = A,T,C or G

<400> 91
tcgagcggcc gnccgggcag gtcccgggtg gttgtttgcc gaaatgggca agttcntnaa 60
ncctgggaag gtggtgcntg tnctggctgg acgctactcc ggacgcnaag ctgtcntcgt 120
gangancatt gat 133

<210> 92
<211> 232
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(232)
<223> n = A,T,C or G

<400> 92
agcgtggtcg cggccgangt ctgtcacttt gcgggggtag cggtcaattc cagccaccag 60
agcatggctg taggggcgat ctgaggtgcc atcatcaatg ttcttcacga tgacaagctt 120
tgcgtccgga gtagcgtcca gccaggacaa gcaccacctt cccacgtntt cangaactng 180
cccatttcgg cataaccacc cgggacctgc ccgggcggnc gctcgaaaag cc 232

<210> 93
<211> 480
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(480)
<223> n = A,T,C or G

<400> 93
agcgtgggtc gcggccgang tctgtangct caccggccag agaagaccac tgtgagcatt 60
ttgccgtata tcctgccctg ccatttgttc acttttttaaa ctaaaatagg aacatccgac 120
acacaccgtt tgcatcgtct tctcccttga tattttaagc attttcccat gtcgtgagtt 180
tctcagaaac atgtttttaa caattgtact atttagtcat ngtccattta ctataattta 240
tctgaccatt tccctactgt taaaatactt aagacggttt ctgatttttc cactatttaa 300
ataatgctgt gatgaatatc tttaaaatct tctgatttct tacttttttc ccccttagat 360
gcctggaagt ggtattttga ggtgaaagag tttgttcatt ttgaanatat ttctgtctct 420
ctctcgacct gatgtgtana cgctcacttc cagttagcag aaccacctta gtttgtgtct 480

<210> 94
<211> 472
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(472)
<223> n = A,T,C or G

<400> 94
tcgagcggnc gcccgggcag ggtctgatgt cantcacaac ttgaagggat gccaatgatg 60
taccaatccn atgtgaaatc tctcctctta tctcctatgc tgganaaggg attacaaagt 120
tatgtggcng ataannaatt ccatgcacct ctantcatcg atgagaatgg agttcatgan 180
ctggtgaacn atggtatctg aacccgatac cangttttgt ttgccacgat angantagct 240
tttatttttg atagaccaac tgtgaaccta ccacacgtct tggacnactg anntctaact 300
atccncaggg ttttattttg cttgttgaac tcttncagct nttgcaaact tcccaagatc 360
canatgactg antttcagat agcattttta tgattcccan ctcattgaag gtcttatnta 420

tntcnttttt tccaagccaa ggagaccatt ggacctcggc cgcgaccacc tn 472

<210> 95
<211> 309
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(309)
<223> n = A,T,C or G

<400> 95
tcgagcggcc gcccgggcag agtgtcgagc cagcgtcgcc gcgatggtgt tgttggagag 60
cgagcagttc ctgacggaac tgaccagact tttccanaag tgccggacgt cgggcancgt 120
ctatatcacc ttgaagaant atgacggtcg aaccaaaccc attccaaaga aangtactgt 180
gganggcttt ganccccgcag acaacnagtg tctgttaaga actaccgatn ggaaanaana 240
anatcagcac tgtgggtgag ctccnaggga agttaataan tttcggatgg gcttattcna 300
acctcctta 309

<210> 96
<211> 371
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(371)
<223> n = A,T,C or G

<400> 96
tcgagcggcc gcccgggcag gtccaccact cacctactcc ccgtctctat agatttgcct 60
gttctgggca gttctcagca atggaatcct actgtgtatc tttttgtgac tggttcttta 120
actcagcatc acattttcaa ggttcatcca tgctgcagcc tggctccgta ctggtgacag 180
tacttcattt ctctctccct tttgttcaga ccaaggtctc cctctgtccc caaggctaaa 240
gtgcagttgg tgtgatcatg gctcactgca gcctcaaact cctggactca aacagtcctc 300
ccatctcagc ctcccaaagt gctgatntta taagttgcaa gccctgcacc cagcctgtat 360
ctccagtttg t 371

<210> 97
<211> 430
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(430)
<223> n = A,T,C or G

<400> 97
tcgancggcc gcccgggcag gtttnttttn tttnttttnt nnnngntagt atttaaagan 60
atttattaaa tcatcttatc accaaaatgg aaacatnttc caactagaaa catgcnacca 120
tcatcttccc cagtccagtc ncaangtcca atattttnct tgcctctgca gataaaaagt 180
tcnnattttt atacccactc ttactccccc ccaaaatttt aattcngtcc tnccctaaaa 240
ttncnccggg taacaantta ccaaaatggc naaccaatta ttttaaanaa aagttgcncn 300

ttnaaaangg aaactttntg gcaanttanc ctcttttccc ttcccacccc ccantttaag 360
gggaaaacaa tggcactttg ctcttgcttn aacccaaaat tgtcttccaa aaactattaa 420
aaatgttnaa 430

<210> 98
<211> 307
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(307)
<223> n = A,T,C or G

<400> 98
tcnaacggcc gcccnggcnn gtctngcngc acctgtgcct canccgtcga tacctggtcg 60
attgggacan ggaanacaat ntggttttca gggaggccac anatttggag aaacggatga 120
attctccttt attccgaant cagctccttg gtctccgtag anggtgatct tgaaattctc 180
ctgttttgaa aactttcttg aanaaacctt acctgctggt tgtatttggt ctcccactcg 240
gacaagtact cgttatccnn ggtactctta atgtgcccac gtnaactccc cgggntggca 300
actggaa 307

<210> 99
<211> 207
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(207)
<223> n = A,T,C or G

<400> 99
gtccnggacc gatgttgcna agannttttct tggtccanta ggttcnaaaa aatgataanc 60
naggtntanc acgtgaagat ntntatanag tcttantnaa aacncntaga tctgnatgac 120
gataantcga anacnggggg aggggntgag gngaggtggn gtganggaag anntgttgat 180
aaaagannna gntgataaga anngagc 207

<210> 100
<211> 200
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(200)
<223> n = A,T,C or G

<400> 100
acntnnacta gaantaacag ncnttctang aacactacca tctgtnttca catgaaatgc 60
cacacacata naaactccaa catcaatttc attgcacaga ctgactgtaa ttaattttgt 120
cacaggaatc tatggactga atctaatgcn nccccaaatg ttgttngttt gcaatntcaa 180
acatnnttat tccancagat 200

<210> 101

<211> 51
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(51)
<223> n = A,T,C or G

<400> 101
tcgagcggcc gcccgggcag gtctgaccag tgganaaatg cccagttatt g 51

<210> 102
<211> 385
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(385)
<223> n = A,T,C or G

<400> 102
aacgtggtcg cggccgaagt ccatggtgct gggattaatc cactgtgacn gtgactctga 60
gttgagttgt ttttcaatct tctccaagcc tgtggactca tcctccacat ccttgggtag 120
taggatgaac atgctgaaga tgctnatttt gaaaaggaac tctatgaatc ttacaattga 180
atactgtcaa tgtttcccca tnacagaacg tggnccccca aggttccatc atctgcactg 240
ggtttgggtg ttctgtcttg gttgactctt gaaaagggac atttcttttt gttttcttga 300
attcanggaa attttcttca tccactttgc ccacaaaagt taggcagcat ttaaccccca 360
anggattttg ggtctgggtc cttcc 385

<210> 103
<211> 189
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(189)
<223> n = A,T,C or G

<400> 103
agcgtggtcg cggccgaagt ctgcagcctg ggactgaccg ggaagctctg attatttacc 60
caccacaggt angttgtgtt ctgaatctca agttcacagg ttaaggctac agcatcctca 120
tcctccacgg ggttggantt gttgctggtg atgaanggtt tggggtggct ctgcataact 180
gttgatctc 189

<210> 104
<211> 181
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(181)

<223> n = A,T,C or G

<400> 104
tcgagcggcc gcccgggcag gtccaggtct ccaccaangc accaccgtgg gaagctggta 60
attgatgccc accttgaagc cnntggggca ccatccncca actggatgct gcgcttggtt 120
ttgatggtgg caatggcaca ttgactcttt tgggaaccac ttcaccacgg tacaacaggc 180
a 181

<210> 105
<211> 327
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(327)
<223> n = A,T,C or G

<400> 105
tcgagcggcc gcccgggcag gtcttctgtg gagtctgcgt gggcatcgtg ggcagtgggg 60
ctgccctggc cgatgctcan aaccccagcc tctttgtaaa gattctcatc gtgganatct 120
ttggcagcgc cattggcctc tttggggtca tcgtcgcaat tcttcanacc tccanaatga 180
anatgggtga ctanataata tgtgtgggtn gggccgtgcc tcacttttat ttattgctgg 240
ttttcctggg acagaactcg ggcgcgaaca cgcttanccg aattccaaca cactggcggg 300
cgttactagt ggatccgagc tcggtac 327

<210> 106
<211> 268
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(268)
<223> n = A,T,C or G

<400> 106
agcgtggtcg cggccgangt ctggcgtgtg ccacatcggt cccacctcgc tttacaaaac 60
agtcctgaac ttnatctaat aaaattattg tacacnacat ttacattaga aaaaganagc 120
tgggtgtang aaaccgggcc tggtgttccc tttaagcgaa ngtggctcca cagttggggc 180
atcgtcgctt cctcnaagca aaaacgccaa tgaaccccna agggggaaaa aggaatgaag 240
gaactgnccn gggangnccg ctccgaaa 268

<210> 107
<211> 353
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(353)
<223> n = A,T,C or G

<400> 107
tcgagcggcc gcccgggcag gtggccaggc catgttatgg gatctcaacg aaggcaaaca 60

cctttacacn ctagatggtg gggacatcat caacgccctg tgcttcagcc ctaaccgcta 120
ctggctgtgt gctgccgcag gccccagcat caagatctgg gatttanagg gaaagatcnt 180
tgtnnatgaa ctgaancnta aattatcagt tccannacca ngcaaaaacc acccngtgca 240
ctccctggcc tggtctgctg atgggacctc gggcgcgaac acgctnancc caattccanc 300
acactgggcg gncgttacta ntggatccga actcnggtac caancttggc gtt 353

<210> 108
<211> 360
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(360)
<223> n = A,T,C or G

<400> 108
agcgtggtcg cggccgaagt cctggcctca catgaccctg ctccagcaac ttgaacagga 60
naagcagcag ctacatcctt aaggtccgga aagttagatg aagatttgga tcctgcattg 120
ncctgcctcc cacctatctc tcccnaatta taaacagcct ccttgggaag cagcagaatt 180
taaaaactct cccnctgccc tnttgaacta cacaccnacc gggaaaacct ttttcanaat 240
ggcacaaaaa tncnagggaa tgcatttcca tgaangaana aactgggtta cccaaaatta 300
ttgggttggg gaaatccngg gggggttttn aaaaaagggc aanccnccaa anaaaaaaac 360

<210> 109
<211> 101
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(101)
<223> n = A,T,C or G

<400> 109
atcgtggtcn cggccgaagt cctgtgtcct ggatgggccg tgtgcancga atccgttggc 60
gactcctaac taccaanaaa angactctcg gaagaaattt c 101

<210> 110
<211> 300
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(300)
<223> n = A,T,C or G

<400> 110
ccanggaaac ccagagtcac atgagatagg gtggctttcg ggacaggggg tcagangaat 60
ggtacatgga tctcagcccc tgatggacac ggaacaggtg tggtcagaac tcccangatt 120
ctgcatccan gatccagtct ctatagaagt tatggatcat tccttcattt cattcccccc 180
ttcatgaaaa aacttctgaa caagcctttt ttctcacttt ggggccctgt ttggcncaag 240
gtnttnantt ggggaaaaaa aaacaaatcc nttccnttan ccctccgtgg ggaatgacct 300

<210> 111
<211> 366
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(366)
<223> n = A,T,C or G

<400> 111
cgagcggccg cccgggcagg tccttgtgtt gccatctgtt ancattgatt tctggaatgg 60
aacanctttc tcaaagtttg gtcttgctan tcatgaagtc atgtcagtgt cttaagtcac 120
tgctgctcac ttccttaccc agggaatata ctgcataagt ttctgaacac ctgttttcan 180
tattcactgt tcctctcctg cccaaaattg gaagggacct catttaaaaa tcaaatttga 240
atcctgaaan aaaaacngga aatntttctc ttggaatttg gaatagaatt attcanttga 300
ataacatgtt ttttcccctt gccttgctct tcncaanaac atctggacct cggccgcgac 360
acctta 366

<210> 112
<211> 405
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(405)
<223> n = A,T,C or G

<400> 112
ctgactncta aacttctaat tcnatcaana taactactct ccttccgtct tncagagtgt 60
tcacaataaa tctgtgaatc tggcatacac agttgctgga aaattgttct tcctccacna 120
aaaggtcaat tgttcnccnc atgaaanaag ataaattgtt catccatcac tnctgaacca 180
tccaaaacgc cggcggaatt attncccccgt tattatgggg aacggaattt tnaataaatt 240
tgggaangaa tggggctttt attgttttgt tttccccctt tcttggcatt gattgggccg 300
caatgggccc cctcgctcan aanntgcccc ggggccggcc gctccaaaac cgaaattccc 360
anccacactt ggcgggccgt tactanttgg atccgaactc ggtta 405

<210> 113
<211> 401
<212> DNA
<213> Homo sapien

<400> 113
ggatagaaga gtatatgggt ttggcaccac ggggtggata ggcaaaacat ttggttgata 60
aggcgcagat tctgaactaa cttgtaaggc ttgtctggtt ttaggacagg taaaatgggg 120
gaatggtaag gagagtttat aggttttagg agcccatgct gtagcaggca agtgataaca 180
ggctttaatc ctttcaaagc atgctgtggg atgagatatt ggcatttgag cggggtaagg 240
gtgattaggt tttaatgaga tggtaagggg tgcatgatcc ggtccgccaa ggaagggaag 300
tagaggtatc ttatacttgt ggggttaagg tgggggggat ataagaggga ggacgccaaa 360
ggaggctttg gattaggaat aaggggcggc aatgagatgc a 401

<210> 114
<211> 401
<212> DNA

<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(401)
<223> n = A,T,C or G

<400> 114
angtccacag gangcangag gccaggctcc gtcccancca gtccatgatg ttgaagagga 60
ggaagcagca catggggttg aagaactgac tccacttccc aggactggtg gagctggtca 120
ccatggctgt ggtggcgggg aagacggaca gggtgacttc tggaagacag tgaagactga 180
aggttttcct ggcttctggg gctcatctgg ctctgattcc ggctccttct ccaggtcaag 240
atccagggtt cagagctact ttcttggggg actactnggg aatcccgttc tcatctgggg 300
gtngaggggg gacggggnaa gggncatgct tgtgacccag gtttcccacc tcggcccgcg 360
accacgctaa ggcccgaatt ncagcacact tggcggcccg t 401

<210> 115
<211> 401
<212> DNA
<213> Homo sapien

<400> 115
atccctgtaa gtctattaaa tgtaaataat acatacttta caacttctct tagtcggccc 60
ttggcagatt aaatctttgc aaaattccat atgtgctatt gaaaaatgaa ataaaacctc 120
agatgtctga attcttattt caaatacagt tatataatta ttttaaatta caatatacaa 180
tttctgttaa atacaactgt taagggattc tgagaacaat tataagatta taataatata 240
tacaaactaa cttctgaaat gacatgggtt gtttccttcc caccctccta ccctctcaaa 300
gagtttttgc atttgctgtt cctggttgca aaaggcaaaa gaaaatctaa aaatagtctg 360
tgtgtgtcca cgacatgctc gctcctttga gaatctcaaa c 401

<210> 116
<211> 301
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(301)
<223> n = A,T,C or G

<400> 116
ngatttaatt gnnagcttct ttttaatgga atnnttggct aaaatgaatt gatgattatg 60
aatatcccta ggaggagtta gcatggannn tgatcatttt cttngnactc ctttangaca 120
nggaaacagg natcagcatg anggtancan aaaccttatn accnangcgc acganctgac 180
ttcttccaaa gagttgnggt tccgggcagc ggtcattgcc gtgcccattg ctggagggct 240
gattctagtg ntgcttatta tgctggccct gaggatgctt ccaanatgaa aataagangc 300
t 301

<210> 117
<211> 383
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature

<222> (1)...(383)
<223> n = A,T,C or G

<400> 117
aattgcaact ggacttttat tgggcagtta cnacaacnaa tgttttcana aaaatatttg 60
gaaaaaatat accacttcat agctaagtct tacagagaan aggatttgct aataaaactt 120
aagttttgaa aattaagatg cnggtanagc ttctgaacta atgcccacag ctccaaggaa 180
nacatgtcct atttagttat tcaaatacca gttgagggca ttgtgattaa gcaaacaata 240
tatttgttan aactttgntt ttaaattact gntncttgac attacttata aaggagnctc 300
taactttcga tttctaaaac tatgtaatac aaaagtatan ntttccccat tttgataaaa 360
gggccnanga tactgantag gaa 383

<210> 118
<211> 301
<212> DNA
<213> Homo sapien

<400> 118
ctgctagaat cactgccgct gtgctttcgt ggaaatgaca gttccttgtt ttttttgttt 60
ctgtttttgt tttacattag tcattggacc acagccattc aggaactacc ccctgcccca 120
caaagaaatg aacagttgta gggagaccca gcagcacctt tcctccacac accttcattt 180
tgaagttcgg gtttttgtgt taagttaatc tgtacattct gtttgccatt gttacttgta 240
ctatacatct gtatatagtg tacggcaaaa gagtattaat ccactatctc tagtgcttga 300
c 301

<210> 119
<211> 401
<212> DNA
<213> Homo sapien

<400> 119
taaggacatg gacccccggc tgattgcatg gaaaggaggg gcagtgttgg cttgtttgga 60
tacaacacag gaactgtgga tttatcagcg agagtggcag cgctttggtg tccgcatgtt 120
acgagagcgg gctgcgtttg tgtggtgaat ggggaggaaa tgtcactgcc gaagaccaaa 180
aacaagcttc ttggtataaa agactcttac agaatatgtg tattgtaatt tattgatctg 240
gatgcttaag tgtcatggac agtaaatgaa tttgaacttt atgtttgagg acatgacatt 300
gggtttgaaa atataaactg cttttgagca gtttaagtca gggcatttga gaataaaata 360
ggaactttct cttcagtttg taaaactctc ttgccctctc t 401

<210> 120
<211> 301
<212> DNA
<213> Homo sapien

<400> 120
tccagagata ccacagtcaa acctggagcc aaaaaggaca caaaggactc tcgacccaaa 60
ctgccccaga ccctctccag aggttggggt gaccaactca tctggactca gacatatgaa 120
gaagctctat ataaatccaa gacaagcaac aaacccttga tgattattca tcacttgggt 180
gagtgcccac acagtcaagc tttaaagaaa gtgtttgctg aaaataaaga aatccagaaa 240
ttggcagagc agtttgtcct cctcaatctg gtttatgaaa caactgacaa acacctttct 300
c 301